RICE UNIVERSITY

An Asymptotic Minimax Analysis of Nonlocal Means on Edges

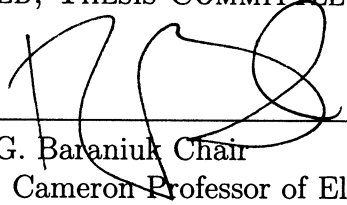
by

Manjari Narayan

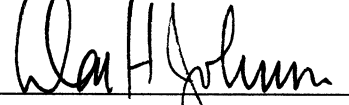
A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Master of Science

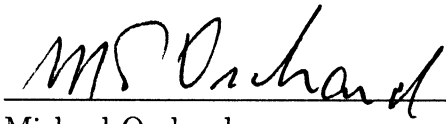
Approved, Thesis Committee:



Richard G. Baraniuk Chair
Victor E. Cameron Professor of Electrical and Computer Engineering



Don H. Johnson
J.S. Abercrombie Professor Emeritus of Electrical & Computer Engineering



Michael Orchard
Professor of Electrical and Computer Engineering

Houston, Texas

December, 2011

ABSTRACT

An Asymptotic Minimax Analysis of Nonlocal Means on Edges

by

Manjari Narayan

This thesis analyzes the non-local means denoising algorithm using the criterion of minimax optimality from statistical decision theory. We show that nonlocal means is minimax suboptimal on images with smooth discontinuities [1] with a rate of convergence of $\mathcal{O}(n^{-1})$ comparable to that of wavelet thresholding. The suboptimality is a consequence of the isotropic nature of the algorithm, and its inability to adapt to the smoothness of the discontinuity. However, all is not lost for nonlocal methods. We also propose an anisotropic nonlocal means algorithm [2] that can attain the optimal rate of $\mathcal{O}(n^{-4/3})$ as well as deliver superior denoising performance using image gradients on synthetic and empirical images, respectively. Nonlocal means is an instance of exemplar based image processing methods. This result broadly implies that exemplar methods that respect anisotropy can yield superior performance in estimating edges in both theory and practice.

Acknowledgements

In order to have something to write, one must have something to say

- PAUL HALMOS

The most wonderful consequence of the past few years in graduate school has been how much I now have to say and the many insights I have to share. Everyone mentioned here, has played a role in my intellectual growth and in uncovering my new found career aspirations. But the greatest debt I owe is to my inimitable advisor, Richard Baraniuk, who has been an amazing role model and mentor. I have learned a great deal from him from the importance of choosing good research problems, to learning how to do research, to the presentation of research. I'm sure I'm not the first to be grateful for the kind of camaraderie he fosters within our research group, but it has certainly enhanced my years at Rice and given me lifelong friends and colleagues.

This work in this thesis has been the result of collaborations with both my advisor and Arian Maleki. I am particularly grateful to have been able to work with Arian, who is both brilliant and prolific, and has helped me learn much of the theoretical machinery in the minimax world. I would also like to thank my committee members, Don Johnson and Mike Orchard for asking me some tough yet useful questions. It made for an exciting and memorable Master's defense. Many thanks to Mike Wakin and Jarvis Haupt for insightful conversations on analyzing nonlocal means.

The ECE department at Rice University fosters a unique *esprit de corps* amongst its students and faculty that has contributed to making graduate school extremely supportive and happy for me. While everyone in the department deserves credit, I do have to thank Rice ECE faculty members, especially Sidney Burrus, Don Johnson

and Behnaam Aazhang who have played such a huge role in shaping the culture of the department.

I can't possibly describe how wonderful it has been to have research buddies like Eva and Jaska. They have been true research siblings and my best friends for the past few years. Many thanks to rest of the DSP group – Chin, Drew, Aswin, Jianing, for being so awesome. I have to mention DSP alumni & Postdocs, Mark (& Kim) , Marco, Mona, Petros, and the rest who have made the group what it is today; y'all are my heroes.

The Valhalla Crew has evolved over the years, many have come and gone; but Gareth, Matt & Amanda, Justin, Sam, Azalia, you guys have been great comrades through the years and have truly made the *Rice Experience* what it is. I definitely miss all of you who have since graduated. Hopefully, I haven't left anyone else out, my apologies and my thanks.

Last but not the least, I am incredibly fortunate in having the parents, grandparents and sister that I have. I love you all.

To my grandparents M.V.V Raman and Kamala Raman

Contents

Illustrations

Tables

Chapter 1

Introduction

Nature loves to hide

-HERACLITUS

1.1 A Brief History of Denoising

Denoising has had a long and varied history in the image processing community. Since the work of Marr [3, 4], we know that visual human perception is sensitive to the presence of boundaries and edges, and thus, edges have been acknowledged as a vital component of images. An important criteria in assessing any image denoising algorithm has therefore been its ability to preserve edges. The fundamental challenge posed by denoising algorithms is that their goal of eliminating noise from an image conflicts with need to preserve useful information in the image. Early denoising algorithms such as linear filtering, assumed that noise was a high frequency component that needed to be filtered out. However, edges and fine textures also constitute high frequency content and they are often eliminated with the noise. This has long motivated denoising algorithms to separate noise from relevant high frequency information more effectively. While state of the art algorithms such as denoising using mixtures of Gaussians to model wavelet coefficients [5] and more recently the BM3D algorithm, based on 3D sparse transforms and nonlocal means filtering [6], have made signifi-

cant strides in this area, edge preservation continues to be the holy grail of image estimation algorithms.

1.1.1 Denoising via Shrinkage

Some of the most successful image denoising methods to date have been wavelet thresholding and shrinkage of other transforms such as wedgelets and curvelets [7–10]. When a dictionary or orthogonal basis represents an image efficiently, most of the energy of the image signal is located in a few transform coefficients. The rest of the zero or nearly zero transform coefficients can be dismissed as noise and thresholded away. One of the disadvantages of these methods is that thus far no known dictionary can efficiently represents all image features. Curvelets, for example, are optimal representations for edges but are not effective for representing textures. The search still continues for a denoising algorithm that is effectively preserves diverse image content.

1.1.2 Denoising via Neighborhood filters

Another widely adopted approach to denoising is neigbhorhood filtering [11, 12]. Here a noisy estimate is replaced by a weighted average of neighboring pixels in the image. The most well known of them, the bilateral filter, chooses weights that are inversely proportional to differences in pixel intensities and pixel locations. The assumption that if two noisy pixels are close to each other radiometrically and spatially, then their true pixel intensities must likewise be close as well, is often violated due to the presence of noise. As a result the bilateral filter suffers from blurry edges.

1.2 Exemplar Methods

The past decade has seen the rise of patch based or exemplar methods in image processing. Inpainting [13, 14] and texture synthesis [15] have been performed using examples or patches from within the image rather than using orthogonal bases or dictionaries. The nonlocal means denoising algorithm is another patch based method where noisy pixels are denoised using information from similar image patches elsewhere in the image. As simple as this denoising method sounds, the idea of leveraging similarity of image patches has been widely adopted to solve many image processing problems including inpainting, super-resolution, and other restoration problems. Peyre [16–18] and Protter, et al. [19] have incorporated nonlocal means inspired constraints in their algorithms, but have also pointed out the lack of theoretical clarity on why nonlocal means works. While we understand a great deal about the best orthogonal bases and dictionaries necessary for approximation and estimation of images, the fundamental limits of performance of exemplar based methods is relatively unknown, particularly on edges.

How well do exemplar based denoising methods perform compared to shrinkage on various transforms? We believe asymptotic minimax analysis is the best tool available to answer this question and that nonlocal means is an ideal representative of exemplar methods to analyze, given its widespread adoption. While our ultimate goal is to identify whether nonlocal means is optimal on smooth regions, textures and edges; in this thesis we make edge analysis our main focus. We first introduce basic concepts and the nonlocal means algorithm in Chapter 2. The results that follow have been adapted from [1, 2] and is the culmination of joint work with Arian Maleki and Richard Baraniuk. In Chapter 3 we derive minimax bounds for the nonlocal means algorithm and demonstrate that it is only as good as shrinkage on wavelets.

In Chapter 4 we prove theoretically, that anisotropy can put nonlocal means on the road to minimax optimality. Finally, in Chapter 5 we demonstrate that an anisotropic variant of nonlocal means outperforms the regular nonlocal means algorithm on edges, precisely because it adapts to edges.

Chapter 2

A Review of Image Estimation

There is an innate risk in computing because to compute is to sample, and one then enters the domain of statistics with all its uncertainties.

- RICHARD W. HAMMING

Consider the following discrete image denoising model where the true pixels of an image located at $i, j \in \{1, 2, \ldots, n\}$ with intensities x_{ij}, have been corrupted independently and identically at each pixel by Gaussian noise $z_{ij} \sim \mathcal{N}(0, \sigma^2)$. The noisy observations of the image are given by

$$y_{ij} = x_{ij} + z_{ij} \quad \forall \quad i, j \in \{1, 2, \ldots, n\} \tag{2.1}$$

Denoising solves the problem estimating the true values of the image x_{ij} from the noisy observations y_{ij}. We shall refer to such estimates as $\hat{x}_{ij}$. In chapter 3, we arrive at the discrete image model in 2.1 by considering each image pixel x_{ij} to be the projection of the true image function f over the pixel basis I_{ij} given by $[\frac{i-1}{n}, \frac{1}{n}] \times [\frac{j-1}{n}, \frac{j}{n}]$ and covers the domain of the image function $[0, 1] \times [0, 1]$. We subsequently use an alternate noise model in chapter 4, where the value of image pixel x_{ij}, given by $f(\frac{i}{n} + u_1, \frac{j}{n} + u_2)$, is obtained by taking samples of the function f at uniformly random shifts u_1, u_2, within an interval $[0, \frac{1}{n}]$ over the pixel domain. This is also known as the jittered grid noise model in [20, 21], where the authors provide minimax analysis of methods under different noise models. Intuitively, it is fair to compare results obtained over both

these image models as the average of the function f over the region I_{ij} is intuitively equivalent to expectation of the the function f sampled at uniformly at random within the interval $(\frac{i}{n} + u_1, \frac{j}{n} + u_2)$. The reason for using such irregular grids for sampling will become apparent over the next few sections, suffice to say that such a sampling scheme is necessary for effective estimation of edges.

2.1 The Minimax Principle

When a function can be described parametrically, such as a sine wave whose frequency ω provides a parametric description, estimating such a function from noisy observations, reduces to estimating the parameter that best describes the function. The best estimator $\hat{f}$ can be found by minimizing the MSE and other parametric estimation methods [22]. However images are high dimensional objects where parametric methods are best avoided. Instead, we prefer to use general classes of functions to describe images so that we can study them using general assumptions such as smoothness or regularity.

In order to analyze estimators of images, where edges are the main feature of interest, we first introduce the Horizon class of functions, before proceeding to key results regarding minimax optimal estimators on such function classes.

2.1.1 Horizon Image Model

A well studied model for images with edges is to consider them as smooth discontinuities piecewise smooth or constant backgrounds was first proposed by Tsybakov and has been extensively used subsequently by Donoho and others. We formally call this the Horizon Image Model Let $H\ddot{o}lder^{\alpha}(C)$ be the class of Hölder functions on $\mathbb{R}$,

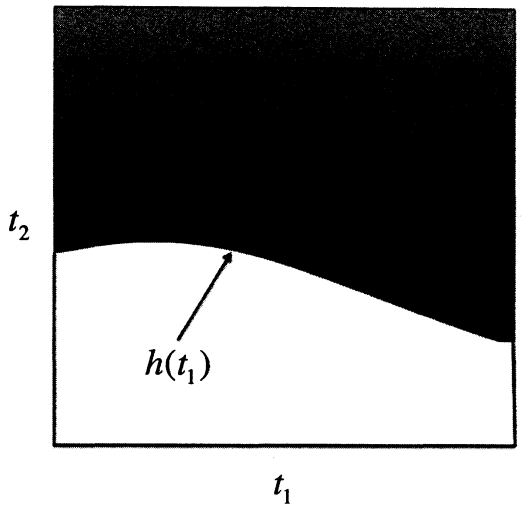


Figure 2.1 : A piecewise constant image where the edge is given by a Horizon function

defined as follows: $h \in H\ddot{o}lder^{\alpha}(C)$ if and only if

$$|h^{(k)}(t_1) - h^{(k)}(t_1')| \leq C|t_1 - t_1'|^{\alpha-k},$$

where $k = \lfloor \alpha \rfloor$. For a one-dimensional function h, we define $f_h : [0,1]^2 \rightarrow \mathbb{R}$ as $f_h(t_1, t_2) = \mathbf{1}_{\{t_2 < h(t_1)\}}$, where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function. Based on this mapping, we define the Horizon class of functions as

$$H^{\alpha}(C) = \{f_h(t_1, t_2) : h \in H\ddot{o}lder^{\alpha}(C) \cap H\ddot{o}lder^{1}(1)\}, \tag{2.2}$$

where α is the smoothness of the edge contour. Figure 2.1 plots a representative function from this class.

2.1.2 Minimax Risk

How then do we find the best possible estimator $\hat{f}$ for a general class of functions where we can no longer optimize over the parameter(s) describing the function ? Minimax optimality is widely accepted a powerful notion from decision theory to

analyze nonparametric estimation methods. We redefine the *risk function* for some general function f and an estimator $\hat{f}$ as

$$R_n(f, \hat{f}) = \mathbb{E}\left(\frac{1}{n^2}\sum_i\sum_j (x_{i,j} - \hat{f}_{i,j})^2\right). \tag{2.3}$$

Alternatively,

$$R_n(f, \hat{f}) = \left(\frac{1}{n^2}\sum_i\sum_j (f_{i,j} - \mathbb{E}\hat{f}_{i,j})^2\right) + \mathbb{E}\left(\frac{1}{n^2}\sum_i\sum_j (\hat{f}_{i,j} - \mathbb{E}\hat{f}_{i,j})^2\right), \tag{2.4}$$

where the first and second terms correspond to the *bias* and *variance* of the estimator $\hat{f}$, respectively.

Let f belong to a class of functions $\mathcal{F}$, e.g., a class of edge-like images that represent edges with different shapes and orientations. The risk defined in (2.3) depends on the specific choice of f. We define the maximal risk of an estimator $\hat{f}$ on the class $\mathcal{F}$ as the risk of the worst-case signal, i.e.,

$$R_n(\mathcal{F}, \hat{f}) = \sup_{f\in\mathcal{F}} R_n(f, \hat{f}).$$

The minimax risk over functions in $\mathcal{F}$ is then defined as the risk of the best possible estimator, i.e.,

$$R_n^*(F) = \inf_{\hat{f}} \sup_{f\in F} R_n(f, \hat{f}).$$

To be precise, we are interested in the asymptotic minimax risk, $\liminf_{n\to\infty} R_n(F)$. This enables us to derive both the theoretically optimal rate of convergence for a given function class as well as compare the rates of convergence of different estimators. Any two estimators with the same convergence rate are therefore considered equivalent in terms of performance. While solving for the exact minimax risk of an estimator is often an intractable problem, the asymptotic rate can often be established through upper and lower bounds. Any two sequences ψ_n^1, ψ_n^2 that are defined within a constant

of each other are equivalent when

$$0 < \lim_{n\to\infty} \inf \frac{\psi_n^1}{\psi_n^2} \leq \lim_{n\to\infty} \sup \frac{\psi_n^1}{\psi_n^2} < \infty$$

Therefore asymptotic minimax upper and lower bounds are considered equivalent when they differ within a constant factor. Also note that given a minimax optimal estimator that converges at the asymptotic rate ψ_n^* and any other estimator that converges at the rate, ψ_n'; if the upper bound $\psi_n' \leq \psi_n^*$ is satisfied, then the estimator with the rate ψ_n' will also be minimax optimal. We shall take advantage of the above insights from [21] in subsequent chapters 3 and 4.

2.1.3 Minimax Optimal Estimators

The minimax rate of convergence on the Horizon class of functions is dependent on the smoothness of the discontinuity in the function class. The following theorem, proved by Korostolev and Tsybakov in [21], specifies the minimax risk of the class of all measurable estimators on $H^\alpha(C)$. The result ensures that no estimator can converge faster than the given rate of $n^{\frac{-2\alpha}{\alpha+1}}$ on the Horizon class. For $\alpha = 2$, the best possible minimax risk is $\mathcal{O}(n^{-4/3})$. Moreover, this minimax rate is only achievable using irregular grid noise models, we introduced earlier. A fixed sampling grid, would restrict the minimax rate for edge estimation to that of the sampling scheme $\mathcal{O}(n^{-1})$.

Theorem. ***(Minimax Risk for Horizon Functions)*** [21] *For $\alpha \geq 1$, the minimax risk of the class $H^\alpha(C)$ is*

$$R_n^*(H^\alpha(C)) \asymp n^{\frac{-2\alpha}{\alpha+1}}. \tag{2.5}$$

Minimax results are available for many transform coding based denoising methods. Donoho and Johnstone [23] [7] have shown that wavelet thresholding has a rate of

$\mathcal{O}(n^{-1})$. Wavelet based methods are unable to efficiently represent two-dimensional discontinuities such as those found in $H^{\alpha}(C)$. However, subsequent work showed that other dictionaries such as wedgelets and curvelets [9, 10] are better suited to represent edges parsimoniously. Many other minimax optimal representations have succeeded curvelets as well. However there has only been a limited understanding of how non-transform based methods perform in the minimax sense. Arias-Castro and Donoho [24] validated the conventional wisdom that median filtering was superior to linear filtering using minimax analysis in low noise conditions. Bilateral filtering [12] is also minimax suboptimal on the Horizon class with a rate of $\mathcal{O}(n^{-2/3})$, no better than linear filtering. The recent popularity of nonlocal means, motivates us to perform a similar analysis.

2.2 Non-Local Means (*NLM*)

Buades, et al. introduced the nonlocal means algorithm [25] where a noisy estimate at (i, j) is replaced by a weighted average of pixels in the image. NLM fundamentally broke with its predecessor the *bilateral filter*, by using weights obtained in the space of image patches rather than image pixels. In NLM the weight associated with a pixel in the averaging operation is inversely proportional to the similarity between the neighborhood of the reference pixel and the one included in the averaging. Thus, the algorithm has no restriction that averaged pixels be radiometrically close to each other, they can be located anywhere in the image so long as the neighborhood resembles that of the reference pixel. Hence the descriptor *non-local*.

$$\hat{f}_{i,j} = \frac{\sum_{(m,\ell)\in S} w_{i,j}(m,\ell) y_{m,\ell}}{\sum_{(m,\ell)\in S} w_{i,j}(m,\ell)}, \tag{2.6}$$

where $S = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$ and $w_{i,j}(m, \ell)$ is set according to the δ_n-neighborhood distance between $y_{i,j}$ and $y_{m,\ell}$. For simplicity , since the reference pixel (i, j) and the algorithm are obvious from the context, we use $w_{m,\ell}$ instead of $w_{i,j}(m, \ell)$.

The weights can be chosen according to a piecewise constant thresholding function or an exponential thresholding function, such as $\exp -\frac{d^2(y_{i,j}, y_{m,\ell})}{h^2}$, which depend on the dissimilarity of the neighborhoods around (i, j) and (m, l). For an image neighborhood of radius δ, the neighborhood distance, $d_\delta(y_{i,j}, y_{m,\ell})$, between two observations is defined as

$$d_\delta^2(y_{i,j}, y_{m,\ell}) = \frac{1}{(2\delta+1)^2} \sum_{n=-\delta}^{\delta} \sum_{p=-\delta}^{\delta} |y_{i+n,j+p} - y_{m+n,\ell+p}|^2 \tag{2.7}$$

It is straightforward to verify that $\mathbb{E}(d_{\delta_n}^2(y_{i,j}, y_{m,\ell})) = d_{\delta_n}^2(x_{i,j}, x_{m,\ell}) + 2\sigma^2$, which suggests the following simplication for setting the weights in our analysis, heretoforth. We know from [1, 25] that the choice of soft or hard thresholding functions for the weights, does not alter the effectiveness of the algorithm either in theory or in practice.

$$w_{i,j}(m, \ell) = \begin{cases} 1 & \text{if } d_\delta^2(y_{i,j}, y_{m,\ell}) \leq 2\sigma^2 + t, \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

where t is the *threshold parameter.*

The primary advantages of the nonlocal means algorithm has been its surprising simplicity, yet competitive performance on diverse image features, particularly edges as well as fine textures. Two important attributes of the nonlocal means estimator, established by Buades, et al., have been the low method noise of the algorithm relative to most contemporary denoising algorithms, as well as its ability to convert noise to noise. The method noise measures the any structured image data that has been discarded as noise, while the ability to convert noise to noise ensures that non local

means does not significantly add irrelevant structure to the image in the process of denoising. We hope that minimax analysis of nonlocal means will complement these relevant empirical attributes of the algorithm from a theoretical perspective and contribute to the development of more sophisticated nonlocal methods for image estimation.

Chapter 3

Suboptimality of Nonlocal Means

The epistemological value of the theory of probability is revealed only by its limit theorems

-Gnedenko and Kolmogorov

Our desideratum is the asymptotic rate of convergence of the nonlocal means algorithm. This will enable us to determine how nonlocal means fares relative to other estimators on edges. While it is often impossible to find the exact minimax estimator, it is generally possible to find asymptotic bounds on the minimax estimator. This chapter presents joint work with Maleki and Baraniuk [1], on the asymptotic minimax upper and lower bounds of the nonlocal means estimator. We will show that the upper bound on the maximal risk of NLM is at best $\mathcal{O}(n^{-1})$, given certain choices in parameters over all functions in the horizon class. Subsequently, we will show that the lower bound of the maximal risk is no worse than $\Theta(n^{-1})$ over the Horizon class, given any choice in the parameters of the algorithm. Appendix A contains detailed lemmas and proofs.

3.1 Oracle Algorithms

The notion of an oracle can simplify the analysis of estimation methods. An oracle gives us the best possible estimate of the unknown signal f assuming access to some typically unknown quantity w. The oracle estimate is therefore better than any

estimator in practice. Formally the oracle $\hat{f_{w^*}}$ minimizes the maximal risk over the class of estimators based on estimates w.

$$R_n(\hat{f}_{w^*}, \mathcal{F}) = \inf_w R_n(\hat{f}_w, \mathcal{F})$$

In the case of NLM, consider an oracle estimator that chooses optimal set of weights w^* based on its access to noise-free weights w, while the real NLM algorithm uses data-driven methods to estimate the weights $\hat{w}$. Any bounds on the performance of the oracle would therefore apply to the performance of NLM as well . This relationship can be formalized in the following way

$$R_n(\hat{f}^{NL}_{\hat{w}}, \mathcal{F}) > R_n(\hat{f}_{w^*}, \mathcal{F}) \tag{3.1}$$

Intuitively, since the oracle estimator contains some knowledge of the true pixels, its lower bound must be smaller than the lower bound of the practical NLM algorithm. Therefore, we shall take advantage of an oracle to find a lower bound on the NLM risk. Before we proceed to define the oracle algorithm, recall the definition of the δ_n-neighborhood distance $d_{\delta_n}(y_{i,j}, y_{m,\ell})$ between two observed neighbhorhoods, defined as,

$$d^2_{\delta_n}(y_{i,j}, y_{m,\ell}) = \frac{1}{\rho_n^2} \sum_{n=-\delta_n}^{\delta_n} \sum_{p=-\delta_n}^{\delta_n} |y_{i+\ell,j+m} - y_{m+n,\ell+p}|^2 - |y_{i,j} - y_{m,\ell}|^2,$$

where $\rho_n^2 = (2\delta_n + 1)^2 - 1$.

We make a small modification above to the usual NLM algorithm of [25], by removing the center element, $|y_{i,j} - y_{n,p}|^2$, from the neighborhood similarity measure. This enables the weights chosen at different pixel locations to be uncorrelated and hence easier to analyze. Since our analysis is primarily asymptotic in nature, i.e. $\delta_n \to \infty$ as $n \to \infty$, the effect of this removal is negligible on the asymptotic rates.

3.1.1 Oracle NLM

Consider a *full oracle* NLM (FNLM) that has access to $\mathbb{E}(d^2_{\delta_n}(y_{i,j}, y_{m,\ell}))$ in setting the weights $w_{m,l}$ in (2.6) and thus sets them using the noise-free values of the pixels,

$$w^F_{i,j}(m,\ell) = \begin{cases} 1 & \text{if } d^2_{\delta_n}(x_{i,j}, x_{m,\ell}) \leq t_{n,}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

In contrast to the *full oracle*, a *semi-oracle* NLM (SNLM) differs much less from the standard NLM by using the following neighborhood distance with partially noise-free pixel values,

$$\bar{d}^2_{\delta_n}(y_{i,j}, y_{m,\ell}) \triangleq \frac{1}{\rho^2_n} \left(\sum_{n=-\delta_n}^{\delta_n} \sum_{p=-\delta_n}^{\delta_n} |x_{i+\ell,j+m} - y_{m+n,\ell+p}|^2 - (x_{i,j} - y_{m,\ell})^2 \right), \tag{3.3}$$

where the weights in (2.6) are set according to

$$w^S_{i,j}(m,\ell) = \begin{cases} 1 & \text{if } \bar{d}^2_{\delta_n}(y_{i,j}, y_{m,\ell}) \leq \sigma^2 + t_{n,}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

Unlike FNLM, SNLM assumes that we only have a noise-free estimates of neighborhood of pixel (i,j), while retaining noisy estimates of the neighborhoods at (m,ℓ) . Therefore, the distances calculated in the SNLM are more accurate than the standard NLM, but a less accurate and more realistic choice relative to the full oracle. In the rest of the chapter, we will use $\hat{f}^N$ and $\hat{f}^S$, to denote the NLM and SNLM estimators, respectively. In next section 3.2 we proceed to obtain an upper bound on the minimax risk. Subsequently in section 3.3we shall find a tight bound on the minimax risk using the oracle.

3.2 Minimax Upper Bound

Our first theorem establishes an *upper bound* on the risk of NLM.

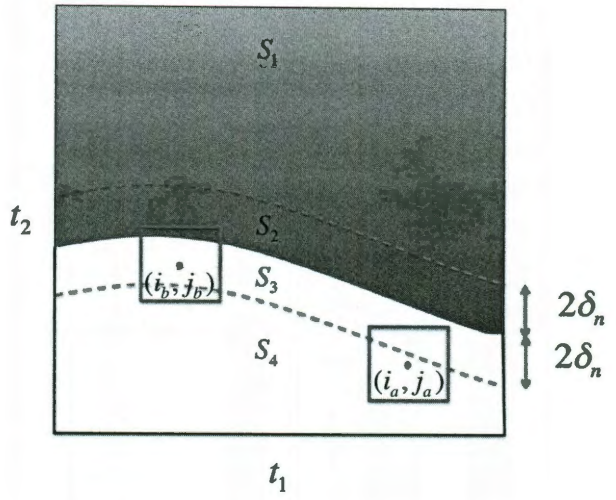


Figure 3.1 : NLM Neighborhoods on a Horizon image. The δ_n-neighborhood of pixel $(i_a, j_a) \in S_4$ does not intersect the edge, while the δ_n-neighborhood of $(i_b, j_b) \in S_3$ intersects with the edge.

Theorem 3.1. *Fix $\epsilon > 0$ and consider NLM denoising with $\delta_n = 2\log^{\frac{1}{2}+\epsilon} n$ and $t_n = \frac{2\sigma^2}{\log^{\frac{\epsilon}{2}} n}$. The risk of this algorithm over the class $H^\alpha(C)$ is*

$$\sup_{f \in H^\alpha(C)} R(f, \hat{f}^N) = O\left(\frac{\log^{\frac{1}{2}+\epsilon} n}{n}\right). \tag{3.5}$$

This bound is within a $\log^3(n)$ factor of the wavelet thresholding. However, this performance is *suboptimal* for $\alpha > 1$. In other words, NLM cannot exploit the C_α smoothness of the edge contour.

3.2.1 Proof Sketch of 3.1

The proof has two main steps. First, we show that the risk of the pixels far from the edge is $O(\log^{3/2+3\epsilon}(n)/n)$. Then, we show that the risk of the pixels whose δ_n neighborhood overlaps with the edge is constant; however there are at most $O(n \log n)$

pixels of the latter kind.

We define a set of partitions as follows. Let $S = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$. For a given horizon function $f_h(t_1, t_2)$, define $S_1 = \{(i,j) \mid \frac{j}{n} > h(\frac{i}{n}) + \frac{\sqrt{2}\delta_n}{n}\}$, $S_2 = \{(i,j) \mid h(\frac{i}{n}) < \frac{j}{n} \leq h(\frac{i}{n}) + \frac{\sqrt{2}\delta_n}{n}\}$, $S_3 = \{(i,j) \mid h(\frac{i}{n}) - \frac{\sqrt{2}\delta_n}{n} \leq \frac{j}{n} \leq h(\frac{i}{n})\}$, and $S_4 = \{(i,j) \mid \frac{j}{n} < h(\frac{i}{n}) - \frac{\sqrt{2}\delta_n}{n}\}$. We use the notation $\sum_{(i,j)\in S_\ell}$ for a double summation over i, j where j satisfies the constraints specified for S_ℓ.

Consider a point $(i,j) \in S_1$. The risk of NLM at this pixel is

$$\mathbb{E}\left(x_{i,j} - \frac{\sum w_{m\ell} y_{m\ell}}{\sum w_{m\ell}}\right)^2,$$

where $x_{i,j} = 0$ since $(i,j) \in S_1$. We refer to the optimal weights chosen given knowledge of the real image to be the oracle weights, formally given by

$$w^\star_{i,j} = \begin{cases} 1 & \text{if } \frac{j}{n} > h(\frac{i}{n}), \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

Let $U = \left(\frac{\sum w_{m\ell} y_{m\ell}}{\sum w_{m\ell}}\right)^2$, and let the event $A = \{w_{m,\ell} = w^\star_{m,\ell},\ \forall (m,\ell) \in S_1 \cup S_4\}$.Then we have

$$\mathbb{E}(U) = \mathbb{E}(U|A)\mathbb{P}(A) + \mathbb{E}(U \mid A^c)\mathbb{P}(A^c) \leq \mathbb{E}(U \mid A)\mathbb{P}(A) + \mathbb{P}(A^c), \tag{3.7}$$

where the last inequality is due to the fact that if the estimate is out of the range of the original signal, then it will be mapped to the closest point in the range. We calculate each term of the above equation separately. Defining $S_{14} = S_1 \cup S_4$ and

$S_{23} = S_2 \cup S_3$ we have,

$$\begin{aligned}
\mathbb{E}(U|A)\mathbb{P}(A) =& \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} y_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell} y_{m\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell}}\right)^2 \mid A\right)\mathbb{P}(A) \\
&\leq \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} y_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell} y_{m\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell}}\right)^2 \\
&\leq \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} x_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell} x_{m\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell}}\right)^2 \qquad (3.8) \\
&\quad + \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} z_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell} z_{m\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}}\right)^2 \qquad (3.9) \\
&\quad + 2\sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} x_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell} x_{m\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell}}\right)^2} \qquad (3.10) \\
&\quad \times \sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_1\cup S_4} w^\star_{m\ell} z_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell} z_{m\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}}\right)^2}.
\end{aligned}$$

The last inequality is due to the Cauchy-Schwartz inequality. To bound $\mathbb{E}(U|A)\mathbb{P}(A)$, we need to bound the two terms (3.8), (3.9) which can be found in Lemma A.3 and Lemma A.4, respectively.

Using Lemma A.3 and Lemma A.4 in Appendix A, we prove that

$$\mathbb{E}(U \mid A)\mathbb{P}(A) = O\left(\frac{\delta_n^3}{n}\right). \qquad (3.11)$$

Finally, using Lemma A.2 and the union bound it is easy to prove that

$$\mathbb{P}(A^c) = O\left(\frac{1}{n^2}\right). \qquad (3.12)$$

Note that the constants in this equality are hidden in the O notation. Since the constants increase with a decrease in ϵ, we cannot set ϵ to 0.

Plugging in (A.5) and (A.4) in Appendix A results in

$$\mathbb{E}\left(x_{i,j} - \frac{\sum w_{m,\ell} y_{m,\ell}}{\sum w_{m,\ell}}\right)^2 = O\left(\frac{\log^{\frac{3}{2}+3\epsilon}(n)}{n}\right) \qquad \forall (i,j) \in S_1.$$

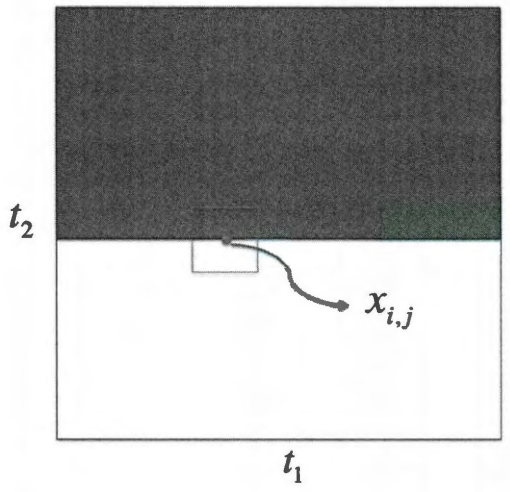


Figure 3.2 : A Simple Horizontal Edge. This image $\mathbf{1}_{\{t_2<0.5\}}$ is used for proving various lower bounds.

Now consider $(i,j) \in S_2 \cup S_3$. In this region we can bound the error by the worst possible risk, which is 1. We will discuss the sharpness of this bound in the next section where we discuss the lower bound for the risk.

Using the bounds provided above for the risks of the pixels in S_1, S_2, S_3 and S_4, we can now calculate the final upper bound for the risk of the NLM as

$$\begin{aligned}
\sup_{f \in H^\alpha(C)} R(f, \hat{f}^{NL}) &= \frac{1}{n^2} \sum_i \sum_j \mathbb{E}(x_{i,j} - \hat{f}^N_{i,j})^2 \\
&\leq \frac{\log^{1+2\epsilon}(n)(|S_1| + |S_4|)}{n^2} + \frac{|S_2| + |S_3|}{n^2} \\
&\leq O\left(\frac{\log^{\frac{1}{2}+\epsilon}(n)}{n}\right).
\end{aligned}$$

In order to derive the last inequality, note that since $h(t_1) \in H\ddot{o}lder^1(1)$, the cardinality of S_2 and S_3 is $O(n\log(n))$. This completes the proof of Theorem 3.1.

3.3 Minimax Lower Bound

The upper bound in 3.1 is for a specific choice of parameters, and so it is natural to ask whether NLM can attain the optimal performance with some other choice of parameters. To answer this question, we consider the lower bound for SNLM, which outperforms standard NLM in general. We make the following mild assumptions:

A_1: The window size $\delta_n \to \infty$ as $n \to \infty$. This assumption is critical to ensuring good performance of NLM.

A_2: The threshold is set to $\sigma^2 + t_n$ as explained in (??) with $t_n > 0$. This ensures that if the neighborhood of pixels around pixel (m, ℓ) is exactly the same as the neighborhood around pixel (i, j), then $w_{m\ell} = 1$ with high probability.

A_3: The threshold t_n is set such that, if the noise-free neighborhoods are different in more than half of their pixels, then $\mathbb{P}(w_{m,\ell} = 1) = o\left(n^{-1}\right)$.

A_4: $\delta_n = O(n^\beta)$ for some $\beta \leq 0.3$.

We derive a lower bound for the risk of SNLM on the image displayed in Figure 3.2. To do so, we consider the pixels just above the edge and prove that the SNLM algorithm has the risk $\Theta(1)$ at these pixels. Since there are $\Theta(n)$ of these pixels, the risk over the entire image is larger than $\Theta(n^{-1})$.

Theorem 3.2. *Suppose that δ_n and t_n satisfy A1–A4. The risk of the SNLM over the class $H^\alpha(C)$ is*

$$\inf_{\delta_n, t_n} \sup_{f \in H^\alpha(C)} R(f, \hat{f}^S) = \Omega(n^{-1}).$$

3.3.1 Proof Sketch of 3.2

Suppose that the parameters of SNLM satisfy assumptions $A1$ -- $A4$. To derive the lower bound we consider the performance of the SNLM algorithm on the simple image displayed in Figure 3.2. For notational simplicity we assume that n is even, and hence all the pixel values are either 0 or 1. The proof will follow four main steps:

1. We consider the pixels that are just above the edge, i.e., $(i, \lceil \frac{n}{2} \rceil)$, and prove that the risk of the NLM on these pixels is lower bounded by a constant that does not depend on n.

2. Using asymptotic arguments we prove that the probability a pixel just below the edge passes the threshold $t_n > 0$ is larger than p_0, where p_0 is a non-zero probability independent of n. Based on this, we use a concentration argument to prove that $\Theta(n)$ of the pixels just below the edge will pass the threshold with high probability. This result is stated in A.2.1.

3. Using symmetry arguments we prove that the probability a pixel that is $\ell < \delta_n/2$ rows* above the edge or below the edge passes the threshold is equal. These results are stated and proved in Lemmas A.5 and A.6 in Appendix A

4. Combining the outcomes of Steps 2 and 3 we show that the risk is minimized if all the edge pixels pass the threshold and the probability that the other pixels pass the threshold is as low as possible. If more zero pixels above the edge pass the threshold, then more pixels with original value 1 will also pass the threshold, and this makes the bias large. Therefore we assume that $p_{n,\ell}$, the probability that a pixel at distance ℓ of the edge passes the threshold, is equal to zero for

*The ℓ^{th} *row* of an image is the set of all pixels of the form (i, ℓ).

$\ell > 1$. However, we have already proven that for $\ell = 1$ the probability is larger than p_0. Theorem 3.2 uses this fact to show that the risk of the NLM estimator is bounded below by a constant, independent of n.

The proof for the lower bound relies heavily on the following proposition which is proved in Appendix A.

Proposition 3.1. *Let $j^* = \lceil \frac{n}{2} \rceil$. For any pixel with coordinates of the form (i^*, j^*), there exists a non-zero constant probability p_0 such that for any δ_n and t_n*

$$\mathbb{P}\left(\sum_m w_{m,j^*-1} - np_0 < -t \right) \leq 4\delta_n e^{-\frac{t^2}{4n\delta_n}}.$$

Corollary 1. *Consider the image displayed in Figure 3.2 and let $\delta_n = O(n^\alpha)$ for $\alpha < 1$. For any δ_n and $t_n > 0$, $\Theta(n)$ of the pixels in J will pass the threshold t_n with very high probability.*

The above corollary holds in a very general setting even if the assumptions A_1 -- A_4 do not hold. In other words, NLM in its most general form is not able to distinguish the pixels right above the edge from the pixels right below the edge. This is due to the fact that the "signal to noise ratio" in the δ_n-neighborhood distance estimates is low at the edge pixels.

To understand why NLM is suboptimal for Horizon class images, consider the estimation of an "edge" pixel value for the particular Horizon class image displayed in Figure 1. An edge pixel (i, j) is a pixel that satisfies $j = \lceil nh(\frac{i}{n}) \rceil$. Intuitively, we expect that a pixel below the edge has a non-zero weight with probability $o(1)$. We proved that this probability is larger than p_0, where p_0 is a constant independent of n. Hence at least $\Theta(n)$ pixels below the edge will be included in the denoised estimate. Since $x_{ij} = 0$, the bias will be larger than $\frac{np_0}{n+np_0+np_0}$. Here np_0 corresponds to the

pixels below the edge that pass the threshold. Clearly the bias will be $\Theta(1)$. This happens due to the low signal to noise ratio in the patch distances near the edge. Combining this with the fact that the distances are symmetric with respect to the edge, we see that the risk on the edge elements is constant. Finally, note that there are $O(n)$ edge pixels and therefore the risk over the entire image is $\Omega\left(n^{-1}\right)$.

This lower bound is characteristic of isotropic neighborhoods used in the weight estimates of NLM. The lower bound is equivalent to that of wavelet thresholding and suboptimal given optimal rate of convergence 2.1.3 for the Horizon class. In next chapter, we shall proceed to examine alternatives to the isotropic neighborhoods used in NLM.

Chapter 4

Presenting Anisotropic Nonlocal Means

Oracle inequalities are neither the beginning nor the end of a theory, but when available are informative tools.

-Iain Johnstone

We now know that the minimax risk of nonlocal means on edges compares to that only of wavelet thresholding. Based on insights from the previous chapter, we propose an anisotropic nonlocal means algorithm (ANLM) that adapts to the smoothness of edges and examine whether it produces minimax optimal results. The work on anisotropic nonlocal means presented in this chapter and the next has been adapted from preliminary results that will appear in [2]. Over the years many anisotropic denoising algorithms have been proposed. Denoising by thresholding in the transform domain, while using anisotropic dictionaries such as curvelets [10, 26], contourlets [27], bandelets [28], and shearlets [29] have been proposed. While wedgelets have been known to be nearly minimax optimal, they are perform poorly on textures. Other popular algorithms such as Perona and Malik's anisotropic diffusion [30, 31] and steerable kernel estimators [32, 33] have also been demonstrably superior to their isotropic counterparts. That anisotropic estimators are uniformly superior to isotropic estimators in the context of nonlocal means, however, has not yet been rigorously defended in the minimax sense. Our ultimate goal is to demonstrate that nonlocal

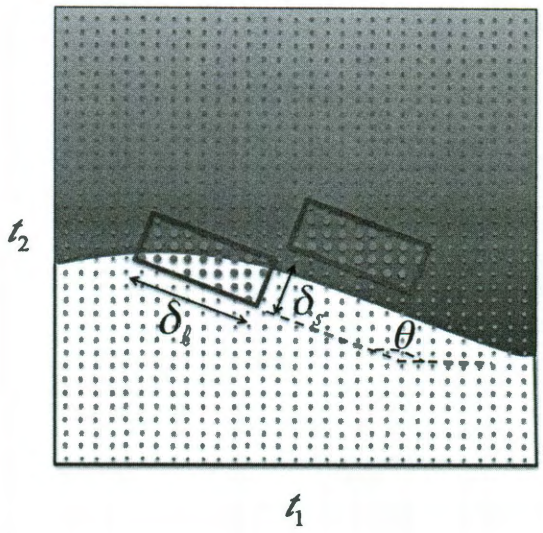


Figure 4.1 : Definition of an Anisotropic Neighborhood.

methods are desirable in their anisotropic form, from both a theoretical and practical perspective. In this chapter, we explore some theoretical results towards this end by analyzing an idealized anisotropic nonlocal means algorithm with access to some information about the true image. A minimax upper bound on such an oracle variant of the algorithm that achieves minimax optimal rates, encourages the possibility of a minimax optimal nonlocal means algorithm in the future. Appendix B contains detailed lemmas and proofs.

4.1 Anisotropic Nonlocal Means *(ANLM)*

We propose an *anisotropic nonlocal means* algorithm that retains the form of the NLM as a weighted average of the image pixels. The novelty is in the 'similarity measure' it exploits in assigning the weights to each pixel. Here we start with the formal definition of the similarity measure. For notational simplicity, suppose the image is sampled

at $\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\} \times \{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\}$ or equivalently the shift variables $u_1 = 0$ and $u_2 = 0$. Let S be $[0,1] \times [0,1]$ and $\tilde{S} = \{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\} \times \{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\}$. Further, let $R^{\theta}_{x,y}(\cdot)$ represent the rotation operator; when applied to a point in $(u,v) \in S$, (u,v) is rotated by θ degrees around the point (x,y), resulting in the transformed point (s,t). For a set S, we define $S_\theta = R^{\theta}_{x,y}(S)$ as

$$(s,t) \in S_\theta \Rightarrow \exists (u,v) \in S \text{ such that } (s,t) = R^{\theta}_{x,y}(u,v).$$

For a given value of θ, δ_s, and δ_ℓ, the $(\theta, \delta_s, \delta_\ell)$-directional neighborhood of the pixel $x_{i,j}$ is defined as

$$I^{i,j}_{\theta,\delta_s,\delta_\ell} = R^{\theta}_{\frac{i}{n},\frac{j}{n}}([\frac{i}{n} - \delta_\ell, \frac{j}{n} + \delta_\ell] \times [\frac{i}{n} - \delta_s, \frac{j}{n} + \delta_s]).$$

$\tilde{I}^{i,j}_{\theta,\delta_s,\delta_\ell} = I^{i,j}_{\theta,\delta_s,\delta_\ell} \cap \tilde{S}$. θ denotes the direction of the neighborhood. δ_s and δ_ℓ ($\delta_s \leq \delta_\ell$) represent the length and the width of the neighborhood. Figure 4.2 displays hypothetical neighborhoods for two different pixels. The discrete $(\theta, \delta_s, \delta_\ell)$-distance between two pixels $y_{i,j}$ and $y_{m,\ell}$ is defined as

$$d^2_{\theta,\delta_s,\delta_\ell}(y_{i,j}, y_{m,\ell}) = \frac{1}{|\mathcal{P}| - 1} \sum_{(p,q) \in \mathcal{P} \setminus \{(0,0)\}} (y_{i+p,j+q} - y_{m+p,\ell+q})^2,$$

where $\mathcal{P} = \{(p,q) \in \mathbb{Z}^2 \mid (\frac{i+p}{n}, \frac{j+q}{n}) \in \tilde{I}^{i,j}_{\theta,\delta_s,\delta_\ell}\}$.

Finally we consider the following estimate for pixel (i,j):

$$\hat{f}^{\theta,\delta_s,\delta_\ell}_{i,j} = \frac{\sum_m \sum_\ell w^{\theta,\delta_\ell,\delta_s}_{i,j}(m,\ell) y_{m\ell}}{\sum_m \sum_\ell w^{\theta,\delta_\ell,\delta_s}_{i,j}(m,\ell)},$$

where the weights are obtained from

$$w^{\theta,\delta_\ell,\delta_s}_{i,j}(m,\ell) = \begin{cases} 1 & \text{if } d^2_{\theta,\delta_s,\delta_\ell}(y_{i,j}, y_{m,\ell}) \leq 2\sigma^2 + t_n, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

This algorithm extends NLM in two ways. First, θ, δ_s and δ_ℓ are free parameters, while in the nonlocal means $\theta = 0$ and $\delta_s = \delta_\ell$. Second, these parameters can

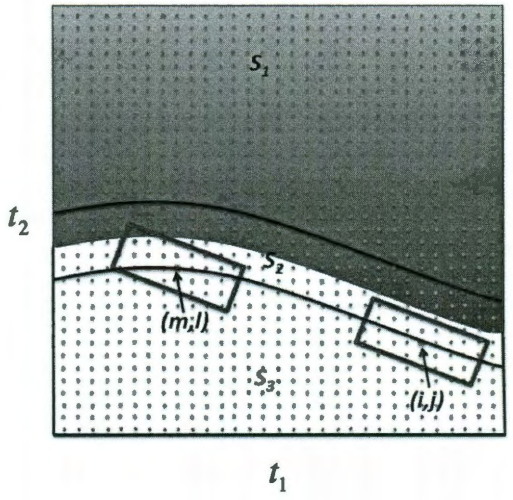


Figure 4.2 : Anisotropic Neighborhood Distance. A hypothetical selection of neighborhoods at pixels (i, j) and (m, ℓ)

be set independently, for the different pixels in the image. Figure 4.3 compare the anisotropic and isotropic neighborhoods on an image with a horizontal linear edge. The anisotropic neighborhood has the following two advantages over the isotropic neighborhood:

- Consider the two pixels $\{y_{i,j_a} \mid j_a = \lceil nh(\frac{i}{n})\rceil\}$ and $\{y_{i,j_b} \mid j_b = \lfloor nh(\frac{i}{n})\rfloor\}$. Figure 4.3 compares the isotropic neighborhood of these two pints with the anisotropic neighborhoods. We assume that the anisotropic neighborhoods are aligned with and elongated in the direction of the edge and they cover the same number of pixels as the isotropic neighborhoods. It is clear from the figure that the anisotropic neighborhoods have larger portion of the edge compared with isotropic neighborhood. Therefore, the probability that (i, j_b) contributes in estimating the pixel relative to (i, j_a) is much lower in ANLM, as opposed to NLM where this probability of contribution was more similar.

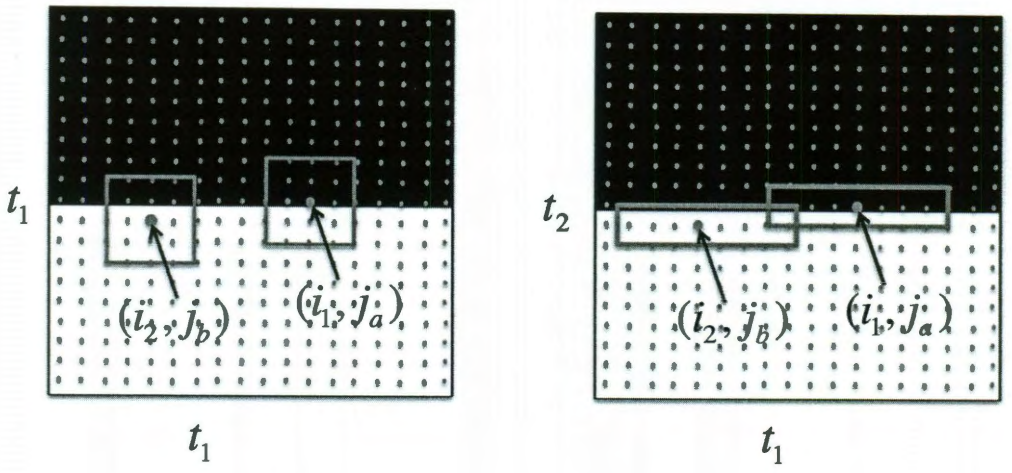


Figure 4.3 : Comparison of Isotropic and Anisotropic Neighborhoods.

- As we show in the theoretical analysis, if image neighborhood does not intersect with the edge, both NLM and ANLM estimate the pixels accurately. However, when anisotropic neighborhoods intersect with the edge, the number of such pixels available for estimation purposes from the correct side of the edge increases. This reduces bias and enables ANLM to provide a higher rate of convergence over NLM.

4.1.1 Empirical Validation of the Benefits of Anisotropy

Figure 4.4 demonstrates that the alignment between the patch and the edge has an impact on the performance. In this example, we have edge with a slope of 135°. We perform anisotropic NLM on a noisy image with a fixed anisotropic patch. The performance of the algorithm is the worst when we use an patch that is orthogonal to the edge, and the best when the patch and the edge are well aligned. The difference in angle results in a change in performance of more 1 dB, and empirically motivates the

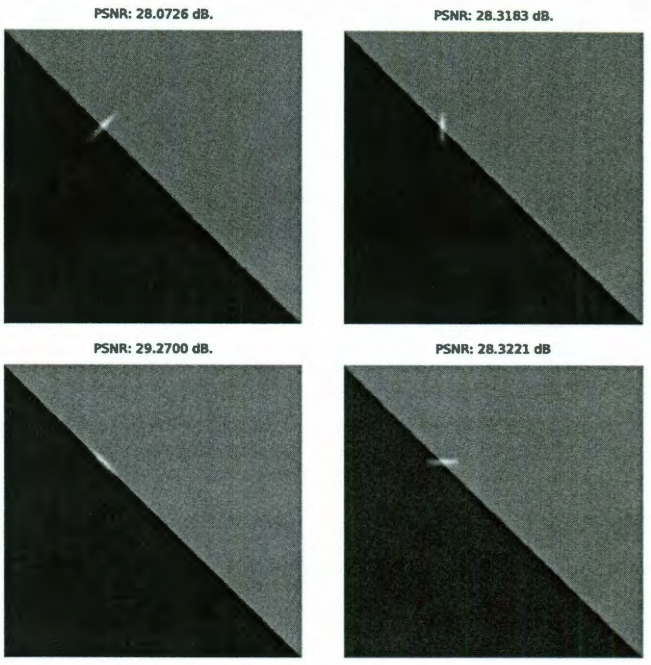


Figure 4.4 : Benefits of Anisotropic Neighborhoods.

need for an algorithm that uses an anisotropic patch of the appropriate orientation, specified by the image geometry. NLM with isotropic neighborhoods outperforms the worst case anisotropic scenario, where the neighborhood is orthogonal to the edge. It however does not perform as well as a well aligned anisotropic neighborhood. We define the *peak signal to noise ratio* (PSNR) in terms of the *mean squared error* (MSE), given by $10\log_{10}\left(\frac{1}{MSE}\right)$ on an image where the pixel intensities have been normalized to the unit interval.

4.2 Oracle ANLM *(OANLM)*

We construct an oracle algorithm that employs the information about the derivative of the edge contour, $h'(t_1)$, in setting the direction of the patches. Let $\Gamma_\gamma = \{(t_1, t_2)$:

$t_1 = \gamma\}$ represent a column of S. Define θ_γ as the angle between the tangent to this contour curve, h, and the horizontal line at $t_1 = \gamma$. The oracle anisotropic nonlocal means is defined as the directional nonlocal means algorithm with the following parameter setting:

- Quadratic scaling: Instead of using square neighborhoods around the points in the image, we consider rectangular neighborhoods of size $\delta_s \times \delta_\ell$. The scaling of δ_s and δ_ℓ depends on the smoothness of the edge. Since we are mainly interested in $\mathcal{H}^2(C)$ we use the quadratic scaling $\delta_s = n^{-4/3}\log^{4/3} n$ and $\delta_\ell = n^{-2/3}\log^{2/3} n$. This scaling is called quadratic since $\delta_s = \delta_\ell^2$. Similar scaling has been used in curvelets and contourlets in a different context.

- Aligned neighborhood: In order to obtain the estimate $\hat{f}_{i,j}$, we set the direction of the neighborhood to $\theta_{i/n}$, i.e., the neighborhood is aligned with the edge. According to this construction, the neighborhood of all the pixels located on $\Gamma_{i/n}$ are aligned.

Figure 4.5 illustrates the optimal neighborhood selection according to OANLM. The following Theorem proved in Chapter B shows that the risk of this algorithm is within a log factor of the minimax rate for a subset of the Horizon class of images, which we denote by $\tilde{\mathcal{H}}^\alpha(C)$.

Theorem 4.1. *Let $\delta_s = n^{-2/3}\log^{2/3} n$ and $\delta_\ell = n^{-4/3}\log^{4/3} n$ in the OANLM algorithm. Also suppose that the algorithm can use the information $h'(t_1)$ and sets θ such that $\tan(\theta) = h'(i/n)$. Then the risk of this algorithm satisfies*

$$R(\tilde{\mathcal{H}}^\alpha(C), \hat{f}^O) = O(n^{-4/3}).$$

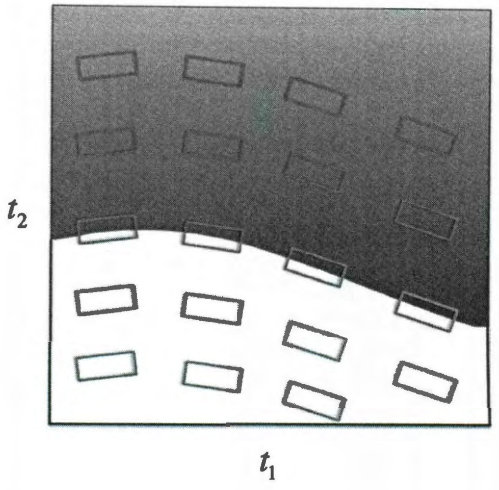


Figure 4.5 : Anisotropic Neighborhoods in the OANLM algorithm. The possible shapes and orientations at at a given point in the image.

4.2.1 Proof Sketch of Theorem 4.1

First, we introduce some notation. For a set $\tilde{A} \subset \tilde{S}$, we define $A \triangleq \tilde{A} \cap S$ and $\bar{A} \triangleq \{(i,j) \mid (i/n, j/n) \in A\}$. Define the following partitions of $\tilde{S}$:

$$\begin{aligned}
\tilde{S}_1 &\triangleq \{(x,y) \mid y > h(x) + (1 + C/2)\delta_s\}, \\
\tilde{S}_2 &\triangleq \{(x,y) \mid h(x) - (1 + C/2)\delta_s \leq y \leq h(x) + (1 + C/2)\delta_s\}, \\
\tilde{S}_3 &\triangleq \{(x,y) \mid y < h(x) - (1 + C/2)\delta_s\}.
\end{aligned}$$

It is important to note that if $(i,j) \in \bar{S}_1$ and $\tan(\theta) = h'(\frac{i}{n})$, then $I^{i,j}_{\theta,\delta_s,\delta_\ell}$ does not overlap with the edge contour. In other words, the correctly aligned neighborhood of (i,j) is always above the edge. The pixels in S_3 also satisfy a similar property. This is clarified in Figure B.1.

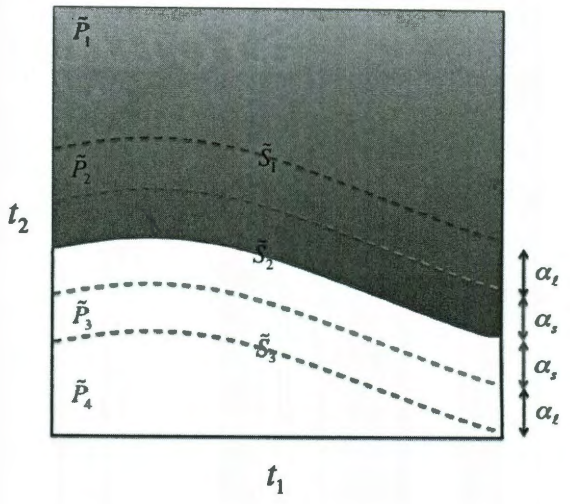


Figure 4.6 : Regions P_1, P_2, P_3, P_4. $\alpha_s = (1 + C/2)\delta_s$ and $\alpha_\ell = 2\delta_\ell - \delta_s$. Every neighborhood of $(i, j) \in P_1$ will lie completely above the edge contour. However, some of the neighborhoods of pixels $(i, j) \in P_2$ may intersect the edge. A similar property holds for regions P_3 and P_4.

We further partition $\tilde{S}_1$ into $\tilde{P}_1$ and $\tilde{P}_2$ and $\tilde{S}_3$ into $\tilde{P}_3$ and $\tilde{P}_4$ such that

$$\begin{aligned}
\tilde{P}_1 &\triangleq \{(t_1, t_2) \mid h(t_1) + (1 + C/2)\delta_s \leq t_2 \leq h(t_1) + 2\delta_\ell + C/2\delta_s\}, \\
\tilde{P}_2 &\triangleq \{(t_1, t_2) \mid h(t_1) + 2\delta_\ell + C/2\delta_s \leq t_2\}, \\
\tilde{P}_3 &\triangleq \{(t_1, t_2) \mid h(t_1) - (1 + C/2)\delta_s \geq t_2 \geq h(t_1) - 2\delta_\ell - C/2\delta_s\}, \\
\tilde{P}_4 &\triangleq \{(t_1, t_2) \mid t_2 \leq h(t_1) - 2\delta_\ell - C/2\delta_s\},
\end{aligned}$$

Any neighborhood of pixel $(i, j) \in P_1$ will lie completely above the edge contour. However, some of the neighborhoods of the pixels $(i, j) \in P_2$ may intersect with the edge. Similarly, this is also true for neighborhoods of pixels $(i, j$ in P_3 and P_4. Figure B.2 displays these regions.

Since the model has random shifts in the pixel locations, u_1, u_2, we have to consider

the following partitions as well. Define $A_2 \triangleq \{(x,y) \in S \mid \exists (u_1,u_2) \in [0,1/n] \times [0,1/n] \text{ s.t. } (x+u_1, y+u_2) \in \tilde{S}_2\}$. $A_1 \triangleq S_1 \backslash A_2$ and $A_3 \triangleq S_3 \backslash A_2$. Also define $AP_2 \triangleq \{(x,y) \in \tilde{P}_2 \mid \exists (u_1,u_2) \in [0,1/n] \times [0,1/n] \text{ s.t. } (x+u_1, y+u_2) \in \tilde{S}_2\}$. $AP_1 = P_1 \backslash AP_2$. Similarly we define AP_3 and AP_4.

We define the oracle weights $w^{\star}_{m,\ell}$ as

$$w^{\star}_{m,\ell} = \begin{cases} 1 & \text{if } x_{m,\ell} = x_{i,j}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

Also define the event $\mathcal{E} = \{w_{m,\ell} = w^{\star}_{m,\ell},\ \forall (m,\ell) \in \overline{AP}_1 \cup \overline{AP}_4\}$. The risk of the OANLM algorithm at (i,j) pixel is then given by

$$\mathbb{E}\left(x_{i,j} - \frac{\sum w_{m,\ell} y_{m,\ell}}{\sum w_{m,\ell}} \right).$$

We need to consider two different distinct cases for pixel (i,j). For each case we calculate the risk and we will then combine the results to obtain an upper bound for the risk of oracle-ANLM.

Case I -- $(i,j) \in \bar{A}_1$: We know that if the anisotropic neighborhood of (i,j), $I^{i,j}_{\theta,\delta_s,\delta_\ell}$, is correctly aligned with the edge contour, i.e., $\tan(\theta) = h'(i/n)$, then it does not intersect with the edge contour. To calculate the ANLM estimate we should first calculate the weights $w_{m,\ell}$. Using Lemmas B.2, B.3 and B.4, we can show that the risk is $O(n^{-4/3})$. Case II -- $(i,j) \in \bar{A}_2$: In this case the random shift is also playing role. If (u_1,u_2) is such that $(i/n+u_1, j/n+u_2) \in S_2$, then the risk is bounded by 1. If (u_1,u_2) is such that $(i/n+u_1, j/n+u_2) \in S_1$, then the risk is $O(n^{-4/3})$ and the proof is similar to the case where $(i,j) \in \bar{A}_1$.

Set $\Omega = \{(u_1,u_2) \in [0,1/n] \times [0,1/n] \mid (i/n+u_1, j/n+u_2) \in S_2\}$. The risk of Pixel

(i,j) is

$$\begin{aligned}\mathbb{E}(U) &= \mathbb{E}(U \mid (u_1,u_2) \in \Omega)\mathbb{P}(\Omega) + \mathbb{E}(U \mid (u_1,u_2) \in \Omega^c)\mathbb{P}(\Omega^c) \\ &\leq n^{-1/3}\log^{2/3}(n) + n^{-4/3} = O(n^{-1/3}\log^{2/3}(n)). \end{aligned} \tag{4.3}$$

Finally

$$\begin{aligned} R_n(f,\hat{f}) &= \frac{1}{n^2}\sum_{(i,j)\in\bar{S}} \mathbb{E}\left(x_{i,j} - \frac{\sum_{(m,\ell)\in\bar{S}} w_{i,j}(m,\ell)y_{m,\ell}}{\sum w_{i,j}(m,\ell)}\right) \\ &= \frac{1}{n^2}\sum_{(i,j)\in\bar{A}_1\cup\bar{A}_3} \mathbb{E}\left(x_{i,j} - \frac{\sum_{(m,\ell)\in\bar{S}} w_{i,j}(m,\ell)y_{m,\ell}}{\sum_{(m,\ell)\in\bar{S}} w_{i,j}(m,\ell)}\right) \\ &\quad + \frac{1}{n^2}\sum_{(i,j)\in\bar{A}_2} \mathbb{E}\left(x_{i,j} - \frac{\sum_{(m,\ell)\in\bar{S}} w_{i,j}(m,\ell)y_{m,\ell}}{\sum_{(m,\ell)\in\bar{S}} w_{i,j}(m,\ell)}\right) \\ &\leq \frac{|\bar{A}_1\cup\bar{A}_2|}{n^2}O(n^{-4/3}\log^2 n) + \frac{|\bar{A}_2|}{n^2}O(n^{-1/3}\log^{2/3} n) \\ &= O(n^{-4/3}\log^2 n). \end{aligned}$$

The last step is the result of Case I and Case II.

Theorem 4.1 is based on a strong oracle assumption that the edge direction is exactly known. Needless to say, such information is not available in practical applications. Hence, we also consider a weaker notion of the oracle. Consider an image $f_h(t_1,t_2) \in H^\alpha(C)$ and let θ_γ be the exact oracle direction at $t_1 = \gamma$. Suppose that the ANLM algorithm has access to an estimate $\hat{\theta}_\gamma$ of θ that satisfies

$$|\hat{\theta}_\gamma - \theta_\gamma| \leq \Theta(n^{-\alpha}). \tag{4.4}$$

The exact edge direction that was assumed to available to the OANLM algorithm corresponds to $\alpha = \infty$ in this model. However, the smaller values of α correspond to the more realistic estimates of the edge orientation. We consider α to be a free parameter. Since we can expect an error in the estimated direction in practice,

$\delta_s = n^{-2/3}\log^{2/3} n$ and $\delta_\ell = n^{-4/3}\log^{4/3} n$ are not necessarily optimal. Consider the extreme case where the edge direction estimator is just a random guess and the error could be as large as $\pi/2$. In this case, isotropic neighborhoods would be preferable to anisotropic ones. The following theorem characterizes the performance of the OANLM algorithm that uses $\hat{\theta}$.

Theorem 4.2. *Consider the OANLM algorithm that uses the direction $\hat{\theta}_\gamma$ satisfying* (4.4). *Set the neighborhood sizes to $\delta_\ell = \min(n^{-2/3}\log^{2/3} n, n^{-1+\alpha/2}\log n)$ and $\delta_s = \log^2 n/(n^2\delta_\ell)$. The risk of the estimator satisfies*

$$R(\tilde{\mathcal{H}}^\alpha(C), \hat{f}^O) = O(\delta_s \mathrm{Poly}(\log n)),$$

where $\mathrm{Poly}(\log n)$ *is a polynomial of degree at most* 2 *in terms of* $\log n$.

If the edge estimate is exact, then the result simplifies to the result of Theorem 4.1. We can now confirm that OANLM algorithm achieves the optimal rate even if there is an error in $\hat{\theta}$ of order $O(\sigma^\beta)$ with $\beta > 2/3$. The proof of this algorithm closely follows that of Theorem 4.1 after redefining the partitions to account for the mismatch in θ and can be found in Appendix B.

In this chapter, we demonstrate that under a discrete image model and oracle information of the edge orientation, an anisotropic nonlocal means algorithm can achieve minimax optimality for a restricted set of Horizon images. In the next chapter, we will demonstrate that an anisotropic implementation of nonlocal means, influenced by the oracle algorithm presented in this chapter, can show superior denoising results in terms of PSNR. It remains for future work to demonstrate that a practical algorithm without access to the oracle can still meet minimax optimal bounds. An extension of these results to the continuous image model will be attempted in the future.

Chapter 5

Performance of Anisotropic Nonlocal Means

It doesn't matter how beautiful your theory is,

it doesn't matter how smart you are.

If it doesn't agree with experiment, it's wrong.

-RICHARD FEYNMAN

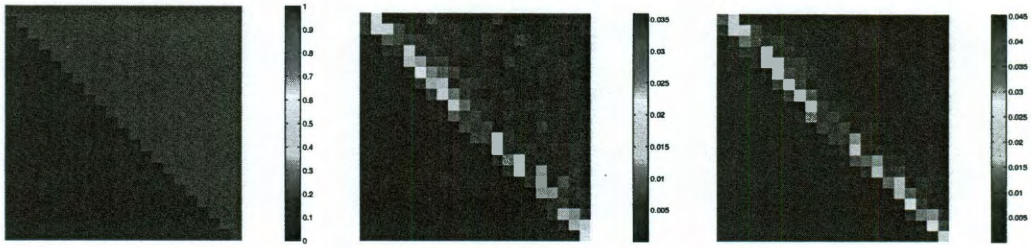
In order for nonlocal means to achieve minimax optimality, the algorithm needs an alternative distance metric capable of discriminating neighborhoods of pixels above and below the edge. In this chapter we propose a practical implementation of ANLM, we substitute the oracle through the use of gradient estimates of the edge orientation. We will demonstrate that through a simple anisotropic implementation, that ANLM is empirically superior to NLM in terms of PSNR values for both synthetic and real test images.

5.1 Analyzing ANLM via Effective Kernels

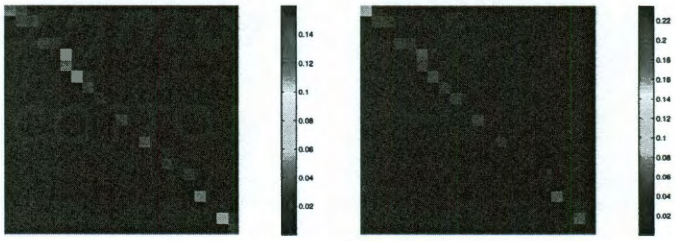
The *effective kernel*, given by

$$K_{ij} = \frac{w_{ij}}{\sum_{S_{ij}} w_{ij}}$$

is the operator used to average pixels in the region S_{ij} to give us the estimate at (i, j). We have found that weights w_{ij} used by the proposed anisotropic distance metric in a restricted neighborhood S_{ij} around pixel (i, j) behave as expected.



(a) Normal Edge (b) Optimal Bias: ANLM (c) Optimal Bias: NLM



(d) Low Bias: ANLM (e) Low Bias: NLM

Figure 5.1 : Effective Kernel Weights of ANLM and NLM.

The more closely the effective kernel resembles an actual edge, the lower the bias of the estimator. So long as the thresholding of the weights is not forced to favor extremely low bias performance i.e. much below $2\sigma^2 + t_n$, we see that ANLM favors asymmetric weighting of pixels above and below the edge while NLM favors them both equally in Figure 5.1. However when the bandwidth of the weights is set to favor extremely low bias, pixels along the edge are not included in the averaging, therefore neither ANLM and NLM effectively denoise the image. The effective kernels thus demonstrate how the averaging operations of ANLM and NLM differ functionally in their choice of weights, with ANLM effectively reaching a superior bias-variance tradeoff.

Our anisotropic algorithm and implementation differ from other anisotropic variants that have been proposed. This can be understood in terms of the effective kernels

shown in Figure 5.1. For instance, Kervrann, et. al [31] adaptively consider isotropic neighborhood sizes in NLM depending on how smoothly the image varies using larger neighborhoods for rapidly changing image features. In [32], the authors introduced an anisotropic algorithm that uses the gradient information to increase the weights of the NLM along the edge and make them smaller across the edges. The modifications in both these methods are akin to modifying the threshold parameter t_n to make NLM assign higher weights to edge like neighborhoods and lower weights to non-edge neighborhoods, as evidenced by the effective kernels in the low bias case in the figure. However, this does not reduce the problem of constant bias, that makes NLM sub-optimal. It also results in an increase in variance along the edge, since the number of pixels that are not thresholded away is much fewer.

5.2 Gradient based Anisotropic Nonlocal Means *(GANLM)*

A practical ANLM algorithm needs to meet two criteria in addition to the theoretical requirements of the oracle in Theorem 4.2 to perform well on natural images – it needs to effectively replace oracle to determine direction of edges and the performance of the algorithm should be robust to errors in estimate of orientation. We propose an ANLM algorithm that mirrors NLM in its computational complexity by using gradients to approximate the tangent to the edge that was previously provided by the oracle. In smooth regions where neighborhoods do not intersect with edges, the anisotropic distance can revert to its isotropic form.

The *gradient based* ANLM (GANLM) comes in two flavors. We demonstrate the efficacy of the gradient approach by estimating the gradients on the original image, so that the gradient estimates are accurate. We call this the oracle GANLM. We also show that an empirical GANLM that estimates the gradients on a noisy image,

wherein these gradient estimates have been obtained from a pilot estimate of the image using NLM, is competitive with the oracle GANLM.

There is a rich literature on robust image gradient estimation [33–35]. To allay any concerns that gradient-based adaptivity is not robust to noise and errors, we recall Theorem 4.2, which establishes the robustness of OANLM to edge angle estimation error. For extremely noisy images, numerous heuristics are possible, including estimating the image gradients for GANLM from an isotropic NLM pilot estimate.

Algorithm 1 *GANLM*: Choosing the best angle

g_{t_1}: Image gradient in t_1 direction

g_{t_2}: Image gradient in t_2 direction

λ: Threshold that determines selection of anisotropic patch.

For every pixel $(i,j) \in I$

$$g_{i,j} = \sqrt{g_{t_1}^2(i,j) + g_{t_2}^2(i,j)}$$
$$\theta_{i,j} = \tan^{-1}\left(\frac{g_{t_1}(i,j)}{g_{t_2}(i,j)}\right)$$

if $g_{i,j} \geq \lambda$ **then**

Perform NLM at pixel (i,j) with $d_{\delta_{\ell_1},\delta_{\ell_2},\theta}$

else

Perform NLM at pixel (i,j) with d_{δ}

end if

5.2.1 Orientation Selection

Consider a $\mathcal{C}^{\in}$ function f which corresponds to the edge in a Horizon image. f_{t_1}, f_{t_2} are tangents to the curve f along the two dimensions t_1 and t_2. The orientation

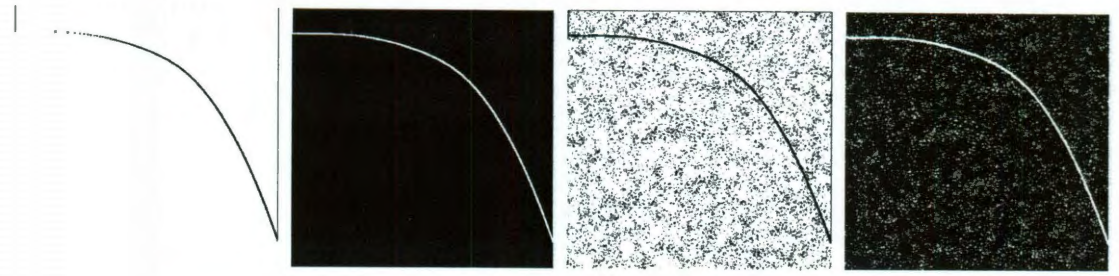
of the tangent to the curve f at a given point would be given by $\tan^{-1}(\frac{f_{t_1}}{f_{t_2}})$. This intuitively suggests that in the absence of noise for a perfect C^2 edge, the ratio of the horizontal and vertical gradients can be used as shown in the pseudocode, to determine the orientation of the edge at any point. Therefore, if $g_{t_1}(i,j)$ and $g_{t_2}(i,j)$ are the estimated image derivatives at pixel (i,j), then we can estimate the local orientation of an edge by $\hat{\theta}(i,j) = \tan^{-1}\left(\frac{g_{t_1}(i,j)}{g_{t_2}(i,j)}\right)$.

5.2.2 Quadratic Scaling

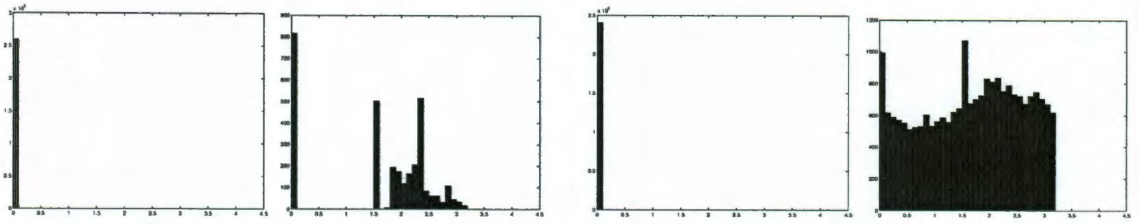
While anisotropic patches improve performance on edges, they may have an unpredictable affect on textures in an image. It is therefore preferable as a control, that the anisotropic algorithm revert to using isotropic patches on non-edge content. We empirically demonstrate that the image gradient enables us to make such a distinction with sufficient accuracy. We choose an anisotropic mask with a fixed quadratic scaling in regions with a large image gradient. Smooth regions or textures give rise to either a non-existent or marginally low image gradient and in such regions, isotropic patches are selected.

5.3 Robustness of GANLM

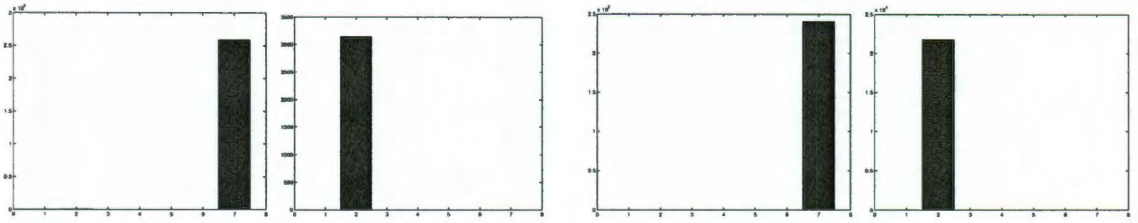
We can see by the performance of the oracle in Figure 5.2 that perfect gradients differentiate edges from the non-edges accurately. While the gradient information from noisy images is not perfect, it definitely mimics the oracle reliably with some false positives i.e. some patches that are not-edges that receive anisotropic neighborhoods. Figures 5.2c show the distribution between the isotropic $7x7$ patches and anisotropic 2×7 patches. The histogram of orientations chosen for the anisotropic neighborhoods, shown in Figures 5.2d, is more uniform for the empirical case due to the contribution



(a) Oracle GANLM (b) Empirical GANLM



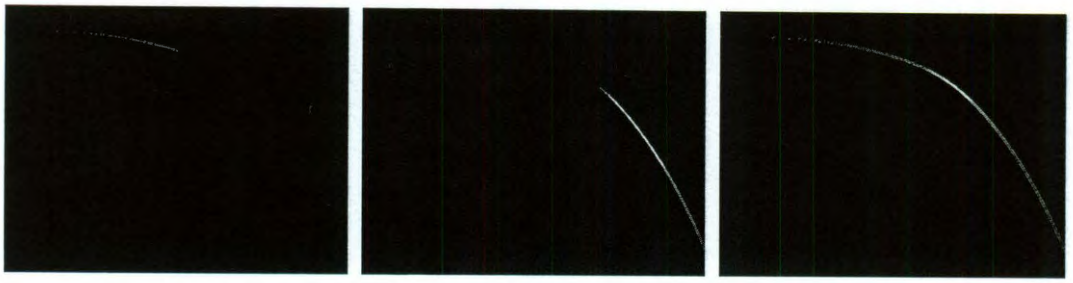
(c) Histogram of Isotropic and Anisotropic Angles



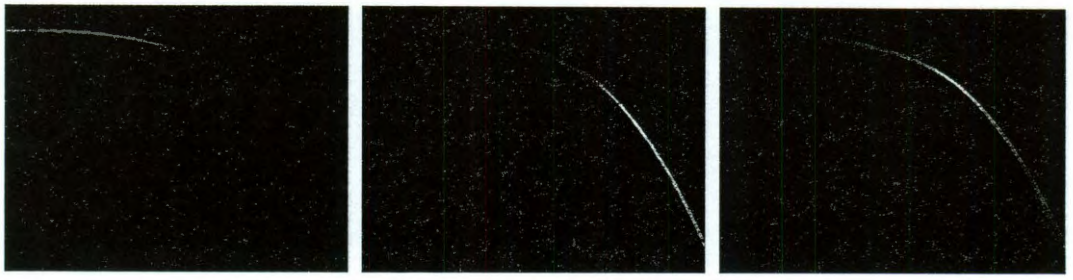
(d) Histogram of Isotropic $\sigma_{l_1} = 7$ and Anisotropic $\sigma_{l_1} = 2$

Figure 5.2 : Isotropic and Anisotropic Patch Selection in Oracle and Empirical GANLM

of the orientations of the false positive anisotropic neighborhoods. However, Fig 5.3 demonstrates that while the oracle and empirical algorithms do not use identical orientations, the error in the orientation estimates falls within an interval of 20°.



(a) Oracle

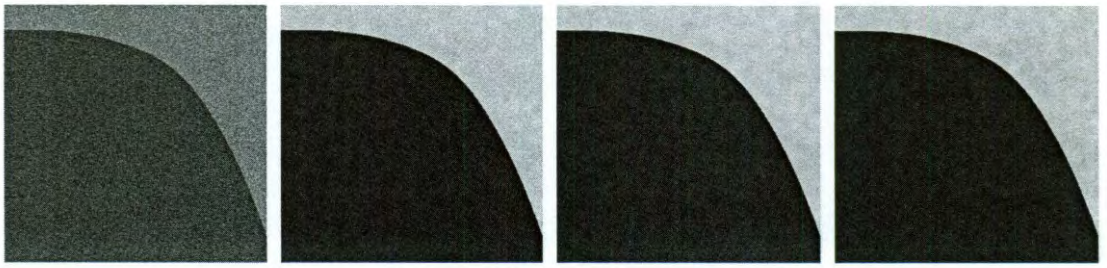


(b) Empirical

Figure 5.3 : Gradient Estimation of Edge Orientation for Angles $\theta = 0°, 120°, 135°$.

5.4 Denoising Performance

Table 5.1 also summarizes the performance of the algorithms introduced in this paper with that of NLM on the natural test images Barbara [36], Boats [37], and Wet Paint [38] and synthetic images at two different noise levels. The performance of the an empirical GANLM algorithm is very close to the oracle GANLM algorithm.Since the oracle GANLM perfectly chooses anisotropic behavior only on edges and not on other image content, it ensures that the gains in PSNR are solely attributable to improvement on edges. The performance on the synthetic horizon images in Table 5.1, we can therefore attribute the 4 dB gains by GANLM over NLM due to improved performance on edges. We have now established that the orientation of the edge, previously provided by an oracle can be well approximated in practice using gradients.



Noise NLM Oracle GANLM Empirical GANLM

Figure 5.4 : GANLM Results on a Horizon image with $\sigma = .25$

Test Image	Algorithm	$\sigma = .25$	$\sigma = .15$
Horizon Class	NLM	30.0881	34.4325
	Oracle GANLM	34.0071	37.9688
	Empirical GANLM	33.4767	37.9533
Simple Edge	NLM	30.4177	34.8934
	Oracle GANLM	34.4787	39.0347
	Empirical GANLM	34.2852	39.0021
Barbara [36]	NLM	22.4784	25.8609
	Oracle GANLM	23.5067	26.6302
	Empirical GANLM	23.5002	26.6046
Wet Paint [38]	NLM	27.6611	30.4863
	Oracle GANLM	29.0206	31.1809
	Empirical GANLM	28.8615	31.0646
Boats [37]	NLM	22.7458	25.8755
	Oracle GANLM	23.8834	26.4898
	Empirical GANLM	23.7525	26.3746

Table 5.1 : Performance of Gradient Anisotropic NLM Algorithms on Test Images

Discussion

Finding the question is often more important than finding the answer.

- JOHN W. TUKEY

We have theoretically confirmed that nonlocal means is competitive with wavelet thresholding with an asymptotic minimax rate of $\mathcal{O}(n^{-1})$. Furthermore, we have shown that an *anisotropic* NLM algorithm can compete with transform shrinkage based methods to achieve the rate of $\mathcal{O}(n^{-4/3})$, under some ideal conditions such as precise knowledge of edge orientations and for a limited range of images in the Horizon class. In the near future, we hope to extend this work with stronger theoretical results on ANLM algorithms as well as efficient and practical alternatives to using gradients as an explicit substitute for oracle information.

NLM has been successfully used in the medical imaging community [39–41]. However, Poisson noise is the relevant noise model for estimation problems in medical imaging. Since the analysis of NLM and ANLM under the standard additive white Gaussian noise model has been promising, we believe avenue for further research would be to extend the analysis of these algorithms to other noise models.

The existence of a minimax optimal NLM algorithms leads us to hope that exemplar methods might be sound alternative to shrinkage methods in the context of image estimation. Thus, one important avenue of future research is to extend theoretical analysis of nonlocal means to other image content, such as periodic textures

and smooth image regions. This would establish the capacity of exemplar methods to adapt to all image content and therefore its universal value.

Appendix A

Proofs for Chapter 3

A.1 Proof of Theorem 3.1

We repeat some results and figures from Chapter 3 in this Appendix for the sake of clarity. Briefly recall the notation we defined in Chapter 3. We divided the image S into a set of partitions S_1, S_2 and S_3 where $S = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$. For a given horizon function $f_h(t_1, t_2)$, define $S_1 = \{(i,j) \mid \frac{j}{n} > h(\frac{i}{n}) + \frac{\sqrt{2}\delta_n}{n}\}$, $S_2 = \{(i,j) \mid h(\frac{i}{n}) < \frac{j}{n} \leq h(\frac{i}{n}) + \frac{\sqrt{2}\delta_n}{n}\}$, $S_3 = \{(i,j) \mid h(\frac{i}{n}) - \frac{\sqrt{2}\delta_n}{n} \leq \frac{j}{n} \leq h(\frac{i}{n})\}$, and $S_4 = \{(i,j) \mid \frac{j}{n} < h(\frac{i}{n}) - \frac{\sqrt{2}\delta_n}{n}\}$.

The NLM risk at a pixel $(i,j) \in S_1$ is given by

$$\mathbb{E}\left(x_{i,j} - \frac{\sum w_{m\ell} y_{m\ell}}{\sum w_{m\ell}}\right)^2,$$

where $x_{i,j} = 0$ since $(i,j) \in S_1$. Also recall that we defined oracle weights given knowledge of the true image as follows,

$$w^{\star}_{i,j} = \begin{cases} 1 & \text{if } \frac{j}{n} > h(\frac{i}{n}), \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

The proof of Theorem 3.1 requires the following lemmas.

A.1.1 Proof of Lemmas A.1 and A.2

Lemma A.1. Let $Z \sim N(0, \sigma^2)$. For $\lambda < \frac{1}{2\sigma^2}$, we have

$$\mathbb{E}(e^{\lambda Z^2}) = \frac{1}{\sqrt{1 - 2\lambda\sigma^2}}.$$

Proof. The proof is a simple integral calculation:

$$\mathbb{E}(\mathrm{e}^{\lambda Z^2}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{(\lambda - \frac{1}{2\sigma^2})Z^2} dZ = \frac{1}{\sigma\sqrt{\frac{1}{\sigma^2} - 2\lambda}}.$$

□

Lemma A.2. Let $Z_1, Z_2, \ldots, Z_n$ be iid $N(0,1)$ random variables. The χ^2_n random variable defined as $\sum_{i=1}^n Z_i^2$ concentrates around its mean with high probability, i.e.,

$$\mathbb{P}\left(\frac{1}{n}\sum_i Z_i^2 - 1 > t\right) \leq \mathrm{e}^{-\frac{n}{2}(t - \ln(1+t))},$$
$$\mathbb{P}\left(\frac{1}{n}\sum_i Z_i^2 - 1 < -t\right) \leq \mathrm{e}^{-\frac{n}{2}(t + \ln(1-t))}.$$

Proof. Here we prove just the first claim; the proof of the second claim follows along very similar lines. From Markov's Inequality, we have

$$\begin{aligned}\mathbb{P}\left(\left(\frac{1}{n}\sum_{i=1}^n Z_i^2\right) - 1 > t\right) &\leq \mathrm{e}^{-\lambda t - \lambda}\mathbb{E}\left(\mathrm{e}^{\frac{\lambda}{n}\sum_{i=1}^n Z_i^2}\right) \\ &= \mathrm{e}^{-\lambda t - \lambda}\left(\mathbb{E}\left(\mathrm{e}^{\frac{\lambda Z_1^2}{n}}\right)\right)^n = \frac{\mathrm{e}^{-\lambda t - \lambda}}{\left(1 - \frac{2\lambda}{n}\right)^{\frac{n}{2}}}. \qquad \text{(A.2)}\end{aligned}$$

The last inequality follows from Lemma A.1. The upper bound proved above holds for any $\lambda < \frac{n}{2}$. To obtain the lowest upper bound we minimize $\frac{\mathrm{e}^{-\lambda t - \lambda}}{\left(1-\frac{2\lambda}{n}\right)^{\frac{n}{2}}}$ over λ. The optimal value of λ is $\lambda^\star = \arg\min_\lambda \frac{\mathrm{e}^{-\lambda t - \lambda}}{\left(1-\frac{2\lambda}{n}\right)^{\frac{n}{2}}} = \frac{nt}{2(t+1)}$. Plugging λ^* into (B.3) proves the result. □

A.1.2 Proof of Lemmas A.3 and A.4

Lemma A.3. Let $w_{m,\ell}$ be the weights of NLM with $\delta_n = \log^{\frac{1}{2}+\epsilon} n$ and $t_n = \frac{2}{\sqrt{\log^{\frac{\epsilon}{2}} n}}$ for $\epsilon > 0$. Also, let $w^\star_{m,\ell}$ be the oracle weights introduced in (B.4). Then

$$\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}}\right)^2 = O\left(\frac{\delta_n^2}{n^2}\right).$$

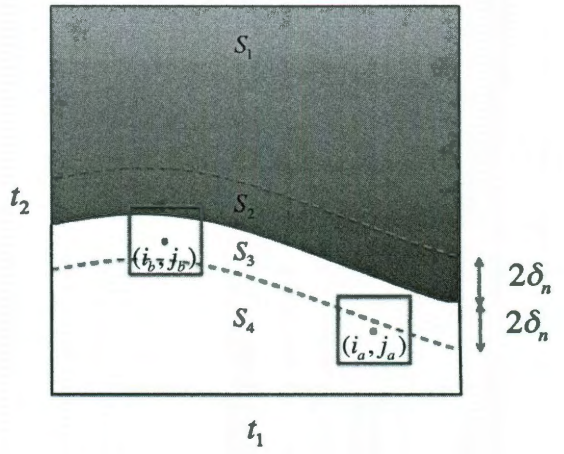


Figure A.1 : The δ_n-neighborhood of pixel $(i_a, j_a) \in S_4$ does not intersect the edge, while the δ_n-neighborhood of pixel $(i_b, j_b) \in S_3$ intersects the edge.

Proof. Define S_f as the set of indices of the pixels whose noise-free value is neither zero nor one. Since the images are chosen from the Horizon class, the cardinality of this set is at most $2n$. Plugging in the values of $x_{m,\ell}$, we have

$$\begin{aligned}
&\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}}\right)^2 \\
&\overset{(a)}{=} \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_3\setminus S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_3\setminus S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_2\setminus S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m\ell}}\right)^2 \\
&\overset{(b)}{\leq} \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_3} 1 + \sum_{(m,\ell)\in S_f} w_{m\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_3} 1 + \sum_{(m,\ell)\in S_f} w_{m,\ell}}\right)^2 \\
&\leq \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_3} 1 + 2n}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_3} 1}\right)^2 = O\left(\frac{\delta_n^2}{n^2}\right),
\end{aligned}$$

where Inequality (b) is due to the fact that the expression after Equality (a) is an increasing function of $\sum_{(m,\ell)\in S_3\setminus S_f} w_{m\ell}$ and a decreasing function of $\sum_{(m,\ell)\in S_2\setminus S_f} w_{m\ell}$. Therefore, we set $w_{m,\ell} = 1$ for $(m,\ell) \in S_3$ and $w_{m,\ell} = 0$ for $(m,\ell) \in S_2$. □

Lemma A.4. Let $w_{m,\ell}$ be the weights of NLM with $\delta_n = \log^{\frac{1}{2}+\epsilon} n$ and $t_n = \frac{2}{\sqrt{\log^{\frac{\epsilon}{2}} n}}$ for $\epsilon > 0$. Also, let $w^\star_{m,\ell}$ be the oracle weights introduced in (B.4). Then we have

$$\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}}\right)^2 = O\left(\frac{1}{n^2}\right).$$

Proof. Since $\sum_{(m,\ell)\in S_{23}} w_{m,\ell} \geq 0$ and we are interested in the upper bound of the risk, we can remove it from the denominator to obtain

$$\begin{aligned}
&\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}}\right)^2\right) \\
\leq\ &\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell}}\right)^2\right) \\
=\ &\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell}}\right)^2\right) + \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell}}\right)^2\right) \\
&+ 2\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell}}\right)\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell}}\right)\right). \qquad \text{(A.3)}
\end{aligned}$$

Since $\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell}}$ is the average of iid random variables, it is not hard to prove that $\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell}}\right)^2 = O(\frac{\sigma^2}{n^2})$. To bound the other two terms in (B.7) we use the notation defined in the last section: $\mathcal{C}^\Delta_{m,\ell} = \{(i,j) : |i-m| < \Delta, |j-\ell| < \Delta\} \cap S$. We also define $\mathbb{E}(\cdot \mid \mathcal{C}^\Delta_{m,\ell})$ as the conditional expectation given the variables in $\mathcal{C}^\Delta_{m,\ell}$. We then have

$$
\begin{aligned}
&\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2 \,\middle|\, \mathcal{C}^{\delta_n}_{i,j}\right)\right) \\
&= \mathbb{E}\left(\frac{\mathbb{E}(\sum_{(m',\ell')\in S_{23}} \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'} \mid \mathcal{C}^{\delta_n}_{i,j})}{(\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell})^2}\right) \\
&= \mathbb{E}\left(\frac{\mathbb{E}(\sum_{(m',\ell')\in \mathcal{C}^{2\delta_n}_{m,\ell}} \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'} \mid \mathcal{C}^{\delta_n}_{i,j})}{(\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell})^2}\right) \\
&= \left(\frac{\sum_{(m',\ell')\in \mathcal{C}^{2\delta_n}_{m,\ell}} \sum_{(m,\ell)\in S_{23}} \mathbb{E}(w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'})}{(\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell})^2}\right) \leq O\left(\frac{\delta_n^3}{n^3}\right).
\end{aligned}
$$

For the last inequality we have used the Cauchy-Schwartz Inequality to prove that $\mathbb{E}(w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'}) \leq 3\sigma^2$. The last term we have to bound in (B.7) is

$$
\begin{aligned}
&\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)\right) \\
&\leq \sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2}\sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2} \\
&\leq O\left(\frac{1}{n^2}\right).
\end{aligned}
$$

This proves the lemma. □

Using Lemma A.3 and Lemma A.4 in (B.6) proves that

$$
\mathbb{E}(U \mid A)\mathbb{P}(A) = O\left(\frac{\delta_n^2}{n^2}\right). \tag{A.4}
$$

Finally, using Lemma A.2 and the union bound it is easy to show that

$$
\mathbb{P}(A^c) = O\left(\frac{1}{n^2}\right). \tag{A.5}
$$

A.2 Proof of Theorem 3.2

A.2.1 Proof of Proposition 3.1

Proposition. *Let $j^* = \lceil \frac{n}{2} \rceil$. For any pixel with coordinates of the form (i^*, j^*), there exists a non-zero constant probability p_0 such that for any δ_n and t_n*

$$\mathbb{P}\left(\sum_m w_{m,j^*-1} - np_0 < -t \right) \leq 4\delta_n \mathrm{e}^{-\frac{t^2}{4n\delta_n}}.$$

Proof. For notational simplicity we use $i = i^*$ and $j = j^*$ in the proof. We have

$$\begin{aligned}
&\mathbb{P}(\bar{d}^2_{\delta_n}(y_{i,j}, y_{m,j-1}) \leq \sigma^2 + t_n) \\
&= \mathbb{P}\left(\frac{1}{\rho_n^2} (\sum_{\ell,p} |x_{i+p,j+\ell} - y_{m+p,j-1+\ell}|^2 - (x_{i,j} - y_{p,j-1})^2) \leq \sigma^2 + t_n \right) \\
&= \mathbb{P}\left(\frac{1}{\rho_n^2} \sum_{\ell,p} (s^2_{\ell,p} - \sigma^2) - \frac{2}{\rho_n^2} \sum_{\ell} s_{\ell,0} \leq -\frac{1}{\rho_n} + t_n \right) \\
&\geq \mathbb{P}\left(\frac{1}{\rho_n^2} \sum_{\ell,p} (s^2_{\ell,p} - \sigma^2) - \frac{2}{\rho_n^2} \sum_{\ell} s_{\ell,0} \leq -\frac{1}{\rho_n} \right),
\end{aligned}$$

where $s_{\ell,m} = z_{m+\ell,j-1+p}$. According to the Berry-Esseen Central Limit Theorem for independent non-identically distributed random variables [42], we know that

$$\mathbb{P}\left(\frac{1}{\rho_n^2} \sum_{\ell} \sum_{p} (s^2_{\ell,p} - \sigma^2) - \frac{2}{\rho_n^2} \sum_{\ell} s_{\ell,0} \leq -\frac{1}{\rho_n} \right) \geq \mathbb{P}(G \leq -1) - \frac{C}{\rho_n},$$

where G is a Gaussian random variable with mean zero and bounded standard deviation. In fact, it is not difficult to confirm that

$$\mathbb{E}(G^2) = 2\sigma^4 + \frac{8\sigma^2\delta_n - 2\sigma^4}{(2\delta_n + 1)^2}.$$

Since $\mathbb{P}(G \leq -1) \geq 2p_0$ ($2p_0$ is $P(G' \leq -1)$ where $G' \sim N(0, 2\sigma^4)$) is non-zero, for large values of n we can ensure that $C/n < p_0$ and therefore that $\mathbb{P}(\bar{d}^2_{\delta_n}(y_{i,j}, y_{m,j-1}) \leq \sigma^2 + t_n) > p_0$. We now prove that even though the weights are correlated, $\Theta(n)$ of

the weights will be equal to 1 with very high probability. Define u_i as $w_{i,j-1}$ and define the process $U = (u_1, \ldots, u_n)$. Break this sequence into $2\delta_n$ subsequences $U_i = (u_i, u_{i+2\delta_n}, u_{i+4\delta_n}, \ldots, u_{n-2\delta_n+i})$. Each U_i has independent and identically distributed elements. Therefore, according to the Hoeffding Inequality, we have $\mathbb{P}(|\sum_{u_j \in U_i} u_j - \frac{n}{2\delta_n}\mathbb{E}(u_i)| > t) \leq 2\mathrm{e}^{\frac{-t^2\delta_n}{n}}$. On the other hand we know that $E(u_i) > p_0$. Therefore,

$$\mathbb{P}\left(\sum_{u_j \in U_i} u_j < \frac{n}{2\delta_n}p_0 - t\right) \leq 2\mathrm{e}^{\frac{-t^2\delta_n}{n}}.$$

Finally we use the union bound to obtain

$$\begin{aligned}\mathbb{P}\left(\sum u_i - np_0 \leq -t\right) &\leq \mathbb{P}\left(\sum_i \sum_{u_j \in U_1} u_j - \frac{n}{2\delta_n}p_0 \leq -t\right) \\ &\leq \mathbb{P}\left(\cup_i\{\omega : \sum_{u_j \in U_i} u_j - \frac{n}{2\delta_n}p_0 \leq -\frac{t}{2\delta_n}\}\right) \leq 4\delta_n \mathrm{e}^{-\frac{t^2}{4n\delta_n}}.\end{aligned}$$

□

Define the set $J = \{(i,j) \mid j = \lfloor jh(\frac{i}{n})\rfloor\}$. It is clear that $|J| = n$. The following Corollary to Proposition A.2.1 shows that NLM sets the weights of most of the pixels in J to 1.

Corollary. Consider the image displayed in Figure 3.2, and let $\delta_n = O(n^\alpha)$ for $\alpha < 1$. For any δ_n and $t_n > 0$, $\Theta(n)$ of the pixels in J will pass the threshold t_n with very high probability.

Proof. Set $t = n^{\frac{3+\alpha}{4}}$ in Proposition A.2.1. □

A.2.2 Proof of Lemmas A.5 and A.6

Lemma A.5. If $|m - i^*| > \delta_n/2$ and $|m' - i^*| > \delta_n/2$, then

$$\mathbb{P}(\bar{d}^2_{\delta_n}(y_{i^*,j^*}, y_{m,j^*-\ell}) \leq \sigma^2 + t_n) = \mathbb{P}(\bar{d}^2_{\delta_n}(y_{i^*,j^*}, y_{m',j^*-\ell}) \leq \sigma^2 + t_n)$$

for any ℓ, m, m'.

The proof of this lemma is obvious and is skipped here.

Lemma A.6. For $\ell < \delta_n/2$,

$$\mathbb{P}(\bar{d}^2_{\delta_n}(y_{i^*,j^*}, y_{m,j^*-\ell}) \leq \sigma^2 + t_n) = \mathbb{P}(\bar{d}^2_{\delta_n}(y_{i^*,j^*}, y_{m,j^*+\ell}) \leq \sigma^2 + t_n).$$

The proof of this lemma is also obvious from symmetry and is skipped here. This concludes the lemmas necessary for the Theorem 3.2 in Chapter 3, which provides a lower bound for the risk of SNLM.

Appendix B

Proofs for Chapter 4

B.1 Proof of Theorem 4.1

We first introduce some notation. For a set $\tilde{A} \subset \tilde{S}$, we define $A \triangleq \tilde{A} \cap S$ and $\bar{A} \triangleq \{(i,j) \mid (i/n, j/n) \in A\}$. Define the following partitions of $\tilde{S}$:

$$\begin{aligned}
\tilde{S}_1 &\triangleq \{(x,y) \mid y > h(x) + (1 + C/2)\delta_s\}, \\
\tilde{S}_2 &\triangleq \{(x,y) \mid h(x) - (1 + C/2)\delta_s \leq y \leq h(x) + (1 + C/2)\delta_s\}, \\
\tilde{S}_3 &\triangleq \{(x,y) \mid y < h(x) - (1 + C/2)\delta_s\}.
\end{aligned}$$

It is important to note that if $(i,j) \in \bar{S}_1$ and $\tan(\theta) = h'(\frac{i}{n})$, then $I^{i,j}_{\theta,\delta_s,\delta_\ell}$ does not overlap with the edge contour. In other words, the correctly aligned neighborhood of (i,j) is always above the edge. The pixels in S_3 also satisfy a similar property. This is clarified in Figure B.1.

We further partition $\tilde{S}_1$ into $\tilde{P}_1$ and $\tilde{P}_2$ and $\tilde{S}_3$ into $\tilde{P}_3$ and $\tilde{P}_4$ such that

$$\begin{aligned}
\tilde{P}_1 &\triangleq \{(t_1,t_2) \mid h(t_1) + (1 + C/2)\delta_s \leq t_2 \leq h(t_1) + 2\delta_\ell + C/2\delta_s\}, \\
\tilde{P}_2 &\triangleq \{(t_1,t_2) \mid h(t_1) + 2\delta_\ell + C/2\delta_s \leq t_2\}, \\
\tilde{P}_3 &\triangleq \{(t_1,t_2) \mid h(t_1) - (1 + C/2)\delta_s \geq t_2 \geq h(t_1) - 2\delta_\ell - C/2\delta_s\}, \\
\tilde{P}_4 &\triangleq \{(t_1,t_2) \mid t_2 \leq h(t_1) - 2\delta_\ell - C/2\delta_s\},
\end{aligned}$$

Any neighborhood of pixel $(i,j) \in P_1$ will lie completely above the edge contour. However, some of the neighborhoods of the pixels $(i,j) \in P_2$ may intersect with

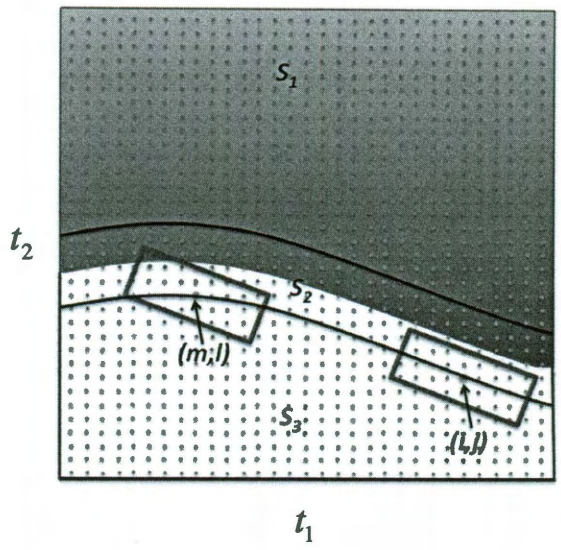


Figure B.1 : Regions S_1, S_2 and S_3. The neighborhood of $(i, j) \in S_3$ is aligned with the edge function, while the neighborhood of (m, ℓ) is incorrectly aligned and may therefore intersect with the edge. Neighborhoods of pixels in S_2 may overlap with the edge even when correctly aligned.

the edge. Similarly, this is also true for neighborhoods of pixels $(i, j$ in P_3 and P_4. Figure B.2 displays these regions. Since the model has random shifts in the pixel locations, u_1, u_2, we have to consider the following partitions as well. Define $A_2 \triangleq \{(x, y) \in S \mid \exists (u_1, u_2) \in [0, 1/n] \times [0, 1/n]$ s.t. $(x+u_1, y+u_2) \in \tilde{S}_2\}$. $A_1 \triangleq S_1 \backslash A_2$ and $A_3 \triangleq S_3 \backslash A_2$. Also define $AP_2 \triangleq \{(x, y) \in \tilde{P}_2 \mid \exists (u_1, u_2) \in [0, 1/n] \times [0, 1/n]$ s.t. $(x + u_1, y + u_2) \in \tilde{S}_2\}$. $AP_1 = P_1 \backslash AP_2$. Similarly we define AP_3 and AP_4.

The following lemma is repeated from Appendix A for convenience.

Lemma B.1. Let $Z_1, Z_2, \ldots, Z_n$ be iid $N(0, 1)$ random variables. The χ^2_n random

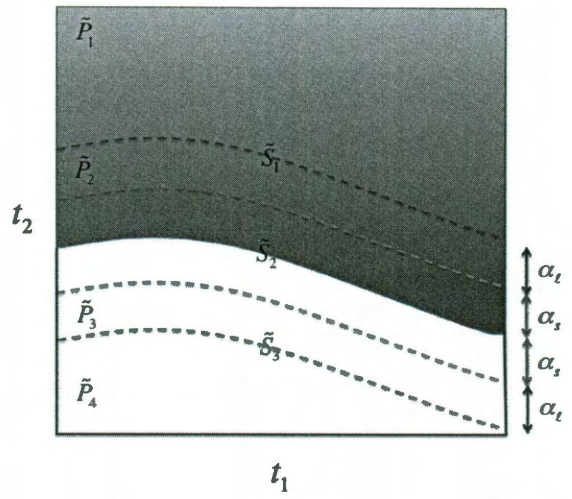


Figure B.2 : Regions P_1, P_2, P_3, P_4. $\alpha_s = (1 + C/2)\delta_s$ and $\alpha_\ell = 2\delta_\ell - \delta_s$. Every neighborhood of $(i, j) \in P_1$ will lie completely above the edge contour. However, some of the neighborhoods of pixels $(i, j) \in P_2$ may intersect the edge. A similar property holds for regions P_3 and P_4.

variable defined as $\sum_{i=1}^{n} Z_i^2$ concentrates around its mean with high probability, i.e.,

$$\mathbb{P}\left(\frac{1}{n}\sum_i Z_i^2 - 1 > t\right) \leq \mathrm{e}^{-\frac{n}{2}(t-\ln(1+t))}, \tag{B.1}$$

$$\mathbb{P}\left(\frac{1}{n}\sum_i Z_i^2 - 1 < -t\right) \leq \mathrm{e}^{-\frac{n}{2}(t+\ln(1-t))}. \tag{B.2}$$

Proof. Here we prove (B.1); the proof of (B.2) follows along very similar lines. From Markov's Inequality, we have

$$\begin{aligned}\mathbb{P}\left(\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2\right) - 1 > t\right) &\leq \mathrm{e}^{-\lambda t-\lambda}\mathbb{E}\left(\mathrm{e}^{\frac{\lambda}{n}\sum_{i=1}^{n} Z_i^2}\right) \\ = \mathrm{e}^{-\lambda t-\lambda}\left(\mathbb{E}\left(\mathrm{e}^{\frac{\lambda Z_1^2}{n}}\right)\right)^n &= \frac{\mathrm{e}^{-\lambda t-\lambda}}{\left(1-\frac{2\lambda}{n}\right)^{\frac{n}{2}}}.\end{aligned} \tag{B.3}$$

The last inequality follows from Lemma A.1 in Appendix A. The upper bound in B.3 holds for any $\lambda < \frac{n}{2}$. To obtain the lowest upper bound we minimize $\frac{\mathrm{e}^{-\lambda t-\lambda}}{\left(1-\frac{2\lambda}{n}\right)^{\frac{n}{2}}}$ over λ. The optimal value of λ is $\lambda^\star = \arg\min_\lambda \frac{\mathrm{e}^{-\lambda t-\lambda}}{\left(1-\frac{2\lambda}{n}\right)^{\frac{n}{2}}} = \frac{nt}{2(t+1)}$. Plugging λ^* into (B.3) proves the result. □

B.1.1 Proof of Lemma B.2

Lemma B.2. Let $\delta_\ell = 2n^{-2/3}\log^{2/3}(n)$, $\delta_s = 4n^{-4/3}\log^{4/3}(n)$, and $t_n = \frac{2\sigma^2}{\sqrt{\log n}}$. Then, for $(m,\ell) \in AP_1$, $\mathbb{P}(w_{m,\ell} = 0) = O(1/n^4)$ and for $(m,\ell) \in AP_4$, $\mathbb{P}(w_{m,\ell} = 1) = O(1/n^4)$.

The proof is completed by plugging in the number of pixels and t_n in Lemma B.1.

Recall that the oracle weights are defined as $w^\star_{m,\ell}$ as

$$w^\star_{m,\ell} = \begin{cases} 1 & \text{if } x_{m,\ell} = x_{i,j}, \\ 0 & \text{otherwise.} \end{cases} \tag{B.4}$$

Also define the event $\mathcal{E} = \{w_{m,\ell} = w^\star_{m,\ell},\ \forall (m,\ell) \in \overline{AP}_1 \cup \overline{AP}_4\}$. The risk of the NLM algorithm at (i,j) pixel is then given by

$$\mathbb{E}\left(x_{i,j} - \frac{\sum w_{m,\ell} y_{m,\ell}}{\sum w_{m,\ell}}\right).$$

Let $U \triangleq \frac{\sum w_{m,\ell} y_{m,\ell}}{\sum w_{m,\ell}}$. We have

$$\mathbb{E}(U) = \mathbb{E}(U \mid \mathcal{E})\mathbb{P}(\mathcal{E}) + \mathbb{E}(U \mid \mathcal{E}^c)\mathbb{P}(\mathcal{E}^c) \le \mathbb{E}(U \mid \mathcal{E})\mathbb{P}(\mathcal{E}) + \mathbb{P}(\mathcal{E}^c). \tag{B.5}$$

Using the union bound and Lemma B.2 it is straightforward to show that $\mathbb{P}(\mathcal{E}^c) = O(1/n^2)$. Define $B = \bar{S}\backslash(\overline{AP}_1 \cup \overline{AP}_4)$ and $\overline{AP}_{14} \triangleq \overline{AP}_1 \cup \overline{AP}_4$. We now calculate each term of (B.5) separately:

$$
\begin{aligned}
&\mathbb{E}(U \mid \mathcal{E})\mathbb{P}(\mathcal{E}) \\
&= \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} y_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell} y_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2 \,\middle|\, \mathcal{E}\right)\mathbb{P}(\mathcal{E}) \\
&\leq \mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} y_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell} y_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2 \\
&\leq \mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2 \\
&\quad + \mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2 \\
&\quad + 2\sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2} \\
&\quad \times \sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2}. \qquad \text{(B.6)}
\end{aligned}
$$

B.1.2 Proof of Lemmas B.3 and B.4

Lemma B.3. Let $w_{m,\ell}$ be the weights of oracle ANLM with $\delta_\ell = 2n^{-2/3}\log^{2/3}(n)$, $\delta_s = 2n^{-4/3}\log^{4/3}(n)$, and $t_n = \frac{2\sigma^2}{\sqrt{\log n}}$. Also, let $w^\star_{m,\ell}$ be the oracle weights introduced in (B.4). Then

$$
\mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2 = O\left(\frac{\log^2 n}{n^{4/3}}\right).
$$

Proof. Define B_1 and B_2 as the set of indices in B for which $x_{m,\ell} = 1$ and $x_{m,\ell} = 0$

respectively. Then we have

$$\begin{aligned}
&\mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}x_{m,\ell}+\sum_{(m,\ell)\in B} w_{m\ell}x_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}+\sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2\\
&\overset{(a)}{=} \mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}x_{m,\ell}+\sum_{(m,\ell)\in B_1} w_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}+\sum_{(m,\ell)\in B_1} w_{m,\ell}+\sum_{(m,\ell)\in B_2} w_{m,\ell}}\right)^2\\
&\overset{(b)}{\leq} \mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}x_{m,\ell}+\sum_{(m,\ell)\in B_1} 1}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}+\sum_{(m,\ell)\in B_1} 1}\right)^2\\
&\leq \mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}x_{m,\ell}+\sum_{(m,\ell)\in B_1} 1}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}+\sum_{(m,\ell)\in B_1} 1}\right)^2 = O\left(\frac{\log^2 n}{n^{4/3}}\right),
\end{aligned}$$

where (b) follows from the fact that the expression on the RHS of (a) is an increasing function of $\sum_{(m,\ell)\in B_1} w_{m\ell}$ and a decreasing function of $\sum_{(m,\ell)\in B_2} w_{m\ell}$. Therefore, we set $w_{m,\ell}=1$ for $(m,\ell)\in B_1$ and $w_{m,\ell}=0$ for $(m,\ell)\in B_2$. Finally it is important to note that $|B_1| = O(n^{4/3}\log^{2/3} n)$, which completes the proof, □

Lemma B.4. Let $w_{m,\ell}$ be the weights of oracle ANLM with $\delta_\ell = 2n^{-2/3}\log^{2/3}(n)$, $\delta_s = 4n^{-4/3}\log^{4/3}(n)$, and $t_n = \frac{2\sigma^2}{\sqrt{\log n}}$. Also, let $w^{\star}_{m,\ell}$ be the oracle weights introduced in (B.4). Then

$$\mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}z_{m,\ell}+\sum_{(m,\ell)\in B} w_{m,\ell}z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}+\sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2 = O\left(\frac{1}{n^2}\right).$$

Proof. Since $\sum_{(m,\ell)\in S_{23}} w_{m,\ell}\geq 0$ and we are interested in the upper bound of the risk,

we can remove it from the denominator to obtain

$$\begin{aligned}
&\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell}}\right)^2\right) \\
&\leq \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}}\right)^2\right) \\
&= \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}}\right)^2\right) + \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}}\right)^2\right) \\
&\quad + 2\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}}\right)\left(\frac{\sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}}\right)\right). \qquad \text{(B.7)}
\end{aligned}$$

It is simple to confirm that

$$\mathbb{E}\left(\frac{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}}\right)^2 = O(\frac{\sigma^2}{n^2}). \qquad \text{(B.8)}$$

Also we define $\mathbb{E}(\cdot \mid \bar{A})$ as the conditional expectation given the variables in $\bar{A}$. We have

$$\begin{aligned}
&\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}}\right)^2\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell}}\right)^2 \,\middle|\, \bar{I}^{\theta,\delta_s,\delta_\ell}_{m,\ell}\right)\right) \\
&= \mathbb{E}\left(\frac{\mathbb{E}(\sum_{(m',\ell')\in B}\sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'} \mid \bar{I}^{\theta,\delta_s,\delta_\ell}_{m,\ell})}{(\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell})^2}\right) \\
&= \mathbb{E}\left(\frac{\mathbb{E}(\sum_{(m',\ell')\in\bar{I}^{\theta,2\delta_s,2\delta_\ell}_{m,\ell}}\sum_{(m,\ell)\in B} w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'} \mid \bar{I}^{\theta,\delta_s,\delta_\ell}_{m,\ell})}{(\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell})^2}\right) \\
&= \mathbb{E}\left(\frac{\sum_{(m',\ell')\in\bar{I}^{\theta,2\delta_s,2\delta_\ell}_{m,\ell}}\sum_{(m,\ell)\in B}\mathbb{E}(w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'})}{(\sum_{(m,\ell)\in\overline{AP}_{14}} w^{\star}_{m,\ell})^2}\right) \leq O\left(\frac{1}{n^2}\right). \qquad \text{(B.9)}
\end{aligned}$$

For the last step we have used the Cauchy-Schwartz inequality to prove that

$$\mathbb{E}(w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'}) \leq 3\sigma^2$$

By using the Cauchy-Schwartz inequality we prove that the last term in (B.7) is also bounded by $O\left(\frac{1}{n^2}\right)$. Finally combining (B.8) and (B.9) completes the proof of Lemma A.4. □

B.2 Proof of Theorem 4.2

We redefine $\tilde{S}_i$ and $\tilde{P}_i$ for this proof. In fact, since there is mismatch between the orientation of the neighborhood and the edge contour, the neighborhood of $\tilde{S}_1$ may intersect with the edge. In order to fix this, we define the new regions called $\tilde{S}_i^\alpha$ and $\tilde{P}_i^\alpha$. If the error in θ is upper bounded by $cn^{-\alpha}$, then define

$$\tilde{S}_1^\alpha = \{(t_1, t_2) | t_2 > h(t_1) + c_\alpha n^{-\alpha}\delta_\ell + \delta_s + C/2\delta_\ell^2\},$$
$$\tilde{S}_3^\alpha = \{(t_1, t_2) | t_2 < h(t_1) - c_\alpha n^{-\alpha}\delta_\ell - \delta_s - C/2\delta_\ell^2\},$$

and $\tilde{S}_2^\alpha = \tilde{S} \backslash (\tilde{S}_1^\alpha \cup \tilde{S}_3^\alpha)$. Furthermore, define $\tilde{P}_1^\alpha = \tilde{P}_1$, $\tilde{P}_4^\alpha = \tilde{P}_2$, $\tilde{P}_2^\alpha = \tilde{S}_1^\alpha \backslash \tilde{P}_1^\alpha$, and $\tilde{P}_3^\alpha = \tilde{S}_3^\alpha \backslash \tilde{P}_4^\alpha$. Using the new partitions instead of the previous partitions the proof is exactly the same as the proof of Theorem ??, and therefore we do not repeat it here.

Bibliography

[1] A. Maleki, M. Narayan, and R. G. Baraniuk, "Minimax suboptimality of nonlocal means on images with sharp edges," *submitted*, 2011.

[2] ——, "Anisotropic nonlocal means denoising," *submitted*, 2011.

[3] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. R. Soc. Lond. B*, vol. 204, no. 1156, pp. 301–328, May 1979.

[4] D. Marr, "Visual information processing: The structure and creation of visual representation," *Proc. R. Soc. Lond. B*, vol. 290, pp. 199–218, 1980.

[5] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans Imag Proc.*, vol. 12, pp. 1338–1351, 2003.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. on Imag. Proc.*, Jan 2007.

[7] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Info. Theo.*, vol. 9, no. 9, pp. 1532–1546, Jan 1995.

[8] D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *Ann. Stat.*, vol. 26, no. 3, pp. 879–921, Jan. 1998.

[9] D. L. Donoho, "Wedgelets: Nearly minimax estimation of edges," *Ann. Stat.*, vol. 27, no. 3, pp. 859 – 897, Jan 1999.

[10] E. Candes and D. L. Donoho, "Curvelets: A Surprisingly Effective Nonadaptive Representation of Objects with Edges," Tech. Rep., 1999.

[11] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[12] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *Proc. Int. Conf. Comp. Vision*, pp. 839–846, Jan. 1998.

[13] J. Aujol and S. Ladjal, "Exemplar-based inpainting from a variational point of view," *Preprint*, Jan. 2009.

[14] A. Wong and J. Orchard, "A nonlocal-means approach to exemplar-based inpainting," *IEEE Int. Conf. Imag. Proc.*, pp. 2600 – 2603, Oct 2008.

[15] E. Levina and P. J. Bickel, "Texture synthesis and nonparametric resampling of random fields," *Annals of Statistics*, vol. 34, no. 4, pp. 1751–1773, 2006.

[16] G. Peyre, "Image processing with non-local spectral bases," *SIAM J. Multiscale Model. and Sim.*, vol. 7, no. 2, pp. 703–730, 2008.

[17] ——, "Sparse modeling of textures," *Jour. of Math. Imag. and Vis.*, Jan 2008.

[18] G. Peyre, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," *Proc. of Eur. Conf. on Comp. Vis.*, Jan 2008.

[19] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the non-local-means to super-resolution reconstruction," *IEEE Trans. on Imag. Proc.*, 2008.

[20] A. Tsybakov, "Multidimensional change-point problems and boundary estimation," *Lecture Notes-Monograph Series*, Jan 1994.

[21] A. Korostelev and A. Tsybakov, *Minimax theory of Image Reconstruction*, ser. Lecture Notes in Statistics. Springer-Verlag, 1993.

[22] E. Candès, "Modern statistical estimation via oracle inequalities," *Acta Numerica*, Jan 2006.

[23] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptopia?" *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 2, pp. 301–369, Jan 1995.

[24] E. Arias-Castro and D. L. Donoho, "Does median filtering truly preserve edges better than linear filtering?" *Ann. Stat.*, vol. 37, no. 3, pp. 1172–1206, Jun 2009.

[25] A. Buades, B. Coll, and J. Morel, "A review of image denoising algorithms, with a new one," *SIAM J. Multiscale Model. and Sim.*, vol. 4, no. 2, pp. 490–530, Jan 2005.

[26] E. J. Candès and D. L. Donoho, "Recovering edges in ill-posed inverse problems: optimality of curvelet frames," *Ann. Stat.*, vol. 30, no. 3, pp. 784–842, Jun 2002.

[27] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[28] G. Peyrè, E. L. Pennec, C. Dossal, and S. Mallat, "Geometrical image estimation with orthogonal bandlet bases," in *Proc. SPIE*, vol. 6701, Aug. 2007, p. 67010M.

[29] G. Kutyniok and D. Labatte, "Construction of regular and irregular shearlet frames," *J. Wavelet Theo. Appl.*, vol. 1, pp. 1–10, 2007.

[30] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629–639, 1990.

[31] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Imag. Proc.*, vol. 7, no. 3, pp. 421 –432, Mar. 1998.

[32] H. Takeda, "Kernel regression for image processing and reconstruction," *MSc thesis, University of California Santa Cruz*, 2006.

[33] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Imag. Proc.*, vol. 16, pp. 349–366, 2007.

[34] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. on Imag. Proc.*, Jan 2006.

[35] A. Wong, P. Fieguth, and D. Clausi, "A perceptually adaptive approach to image denoising using anisotropic non-local means," *IEEE Intern. Conf. on Imag. Proc.*, pp. 537 – 540, Sep 2008.

[36] X. Feng and P. Milanfar, "Multiscale principal components analysis for image local orientation estimation," in *Asilomar Conference on Signals, Systems and Computers*, vol. 1, nov. 2002, pp. 478 – 482 vol.1.

[37] X. Li and M. Orchard, "Edge-directed prediction for lossless compression of natural images," *IEEE Transactions on Image Processing,*, vol. 10, no. 6, pp. 813 –817, jun 2001.

[38] J. Bigun, G. Granlund, and J. Wiklund, "Multidimensional orientation estimation with applications to texture analysis and optical flow," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 8, pp. 775 –790, Aug. 1991.

[39] Unknown. (2011) Barbara test image. http://decsai.ugr.es/~javier/denoise/test_images/index.htm. [Online]. Available: http://decsai.ugr.es/~javier/denoise/test_images/index.htm

[40] ——, "Boats test image," USC-SIPI Image Database, 2011.

[41] (2011) Wet-paint test image. http://www.ece.rice.edu/~wakin/images/. [Online]. Available: "http://www.ece.rice.edu/~wakin/images/"

[42] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 425 – 441, Apr 2008.

[43] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot, "Nonlocal means-based speckle filtering for ultrasound images," *IEEE Trans. Imag. Proc.*, vol. 18, no. 10, pp. 2221 – 2229, Oct 2009.

[44] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salamero, "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 2, pp. 442 –454, feb. 2010.

[45] C. Stein, *Approximate computation of expectation.* Institute of Mathematical Statistics, 1986.